

Mining for Gold: Text Analytics in Insurance

Liam McGrath, ACAS, Willis Towers Watson

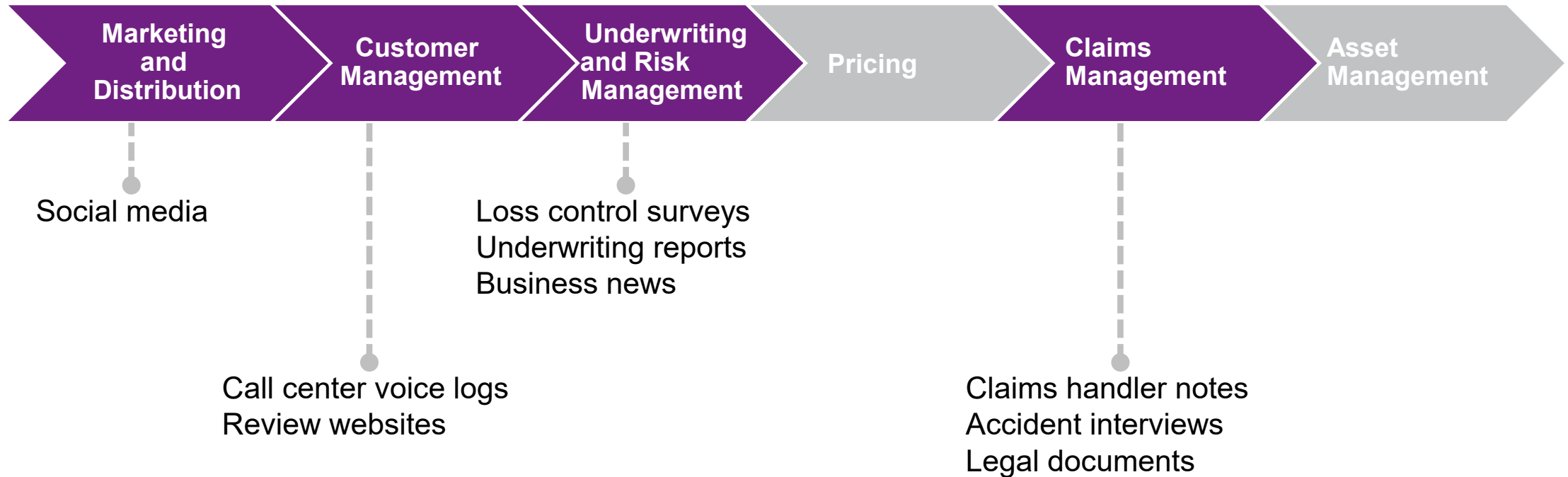
Yelena Kropivnitskaya, PhD, The Wawanesa Mutual Insurance Company

Agenda

- Natural Language Processing – why should you care?
- How do we structure unstructured data?
- Text mining for feature engineering
- Case studies
 - Underwriting
 - Claims – Independent Medical Examination
 - Claims – At-fault rating
- Conclusions

Sources of unstructured text data

The insurer's goldmine



Tapping into these sources allows us to...

Make the most of data we already have

Insurers already collect various types of text data through normal business operations. Text mining ensures it isn't collecting dust!

Fill gaps in structured data

Structured data has limitations. Text data can provide more nuance to fill in gaps.

Quantify internal knowledge consistently

Machine learning can quantify the implicit knowledge of adjusters and underwriters while also smoothing out the natural variation of human decision makers.

Polling question

QUESTION: How would you categorize your company's experience so far using unstructured data?

ANSWERS:

Don't currently use

Tried it but haven't seen a benefit

Seeing a benefit but with room for improvement

Getting the most out of this data source

Expected benefits have yet to be realized

Personal Lines		Expected for 2019 (in 2017)	Actual for 2019	Expected for 2021
	Unstructured internal claim information	66%	38%	69%
	Unstructured internal underwriting information	50%	18%	67%

Commercial Lines		Expected for 2019 (in 2017)	Actual for 2019	Expected for 2021
	Unstructured internal claim information	91%	53%	81%
	Unstructured internal underwriting information	63%	16%	66%

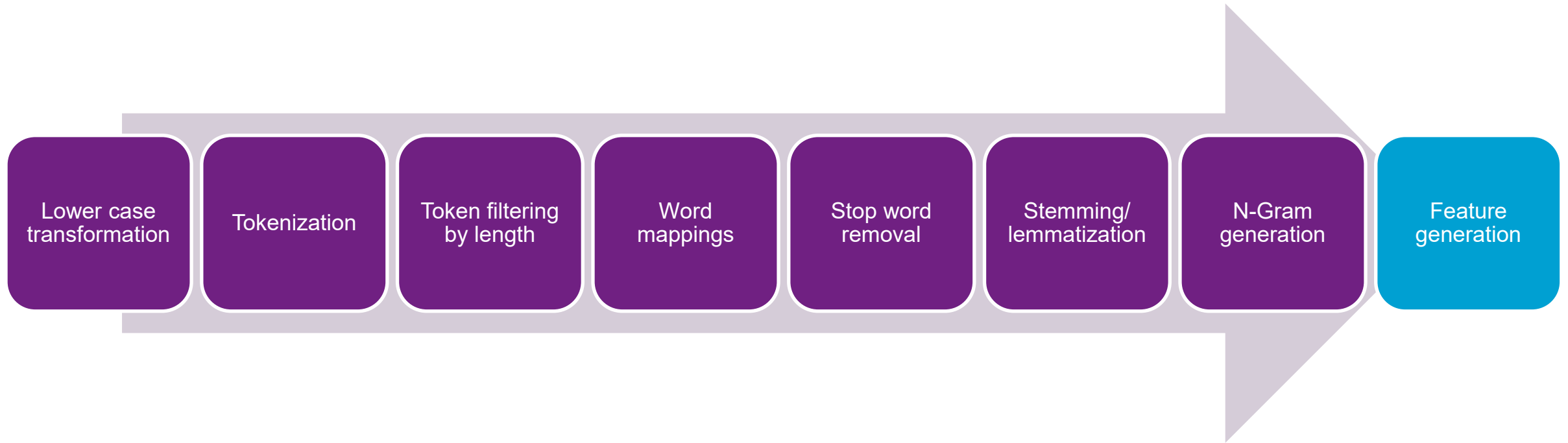
Natural Language Processing will allow us to extract much more out of unstructured data

Example: a claim note

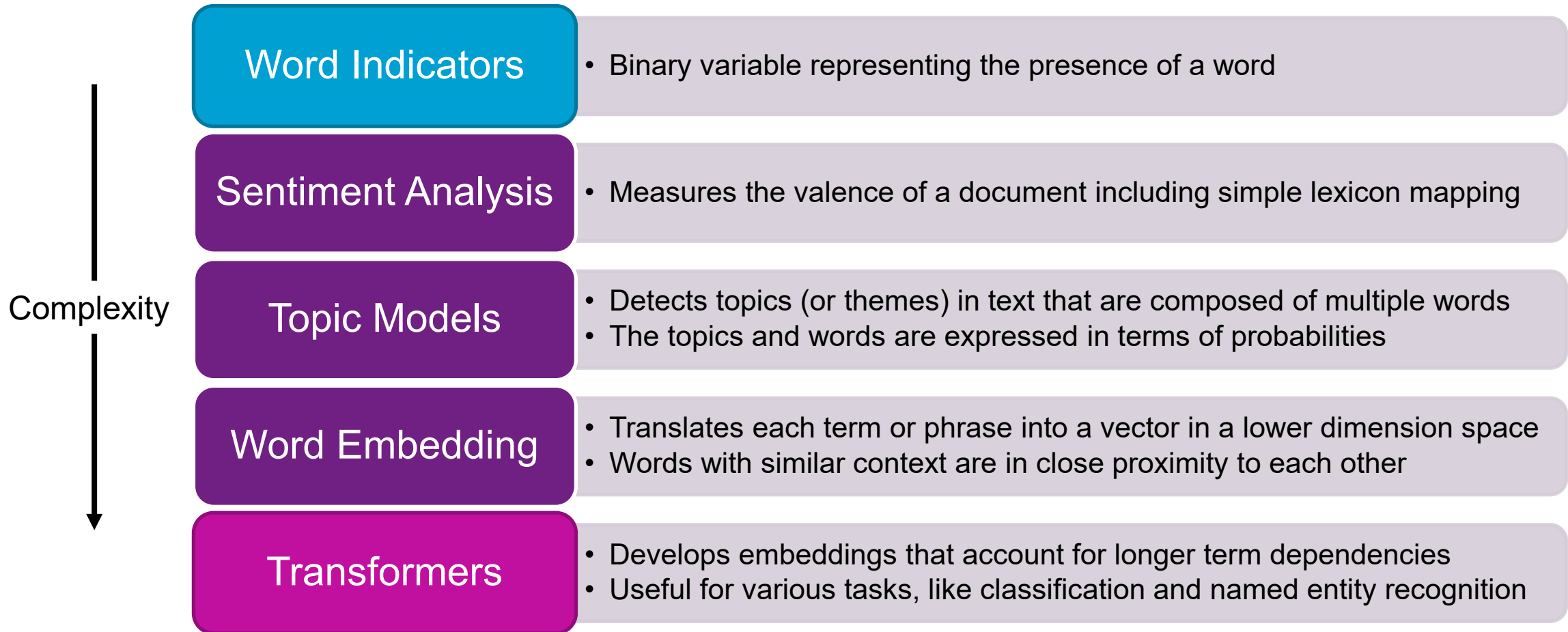
PC to Jane Doe/insd: DOI: 01/01/16 Clmt was carrying drywall up steps with a co-worker. When the co-worker reached the top of the stair steps, he started to walk faster, causing clmt to twist his back and strain his R/shoulder. C/o pain in mid-back & R/shoulder Incident witnessed by co-worker, John Smith Clmt did not report back sprain injury to his supervisor until D/L NLT...

- Problems with unstructured data
 - Junk words, numbers, and formatting
 - Many meanings for a word (polysemy)
 - Many words with the same meaning (synonymy)
 - Negation
 - Abbreviations and acronyms

Text processing



Feature engineering



Case study: Commercial lines underwriting



Goal

- Segment risks using underwriting reports

Features

- Structured fields typically used for rating and underwriting: policy details, exposure information, loss history, 3rd party data
- Underwriting reports, loss descriptions, and loss control surveys
 - Topic Modeling

What do topics tell us about a document?

Example topics:

Topic 1 Top Words:

farm, crops, tractors, harvesting, acres, plant, grow

Topic 2 Top Words:

drug, testing, checks, required, employment, mvr, physical

Topic 3 Top Words:

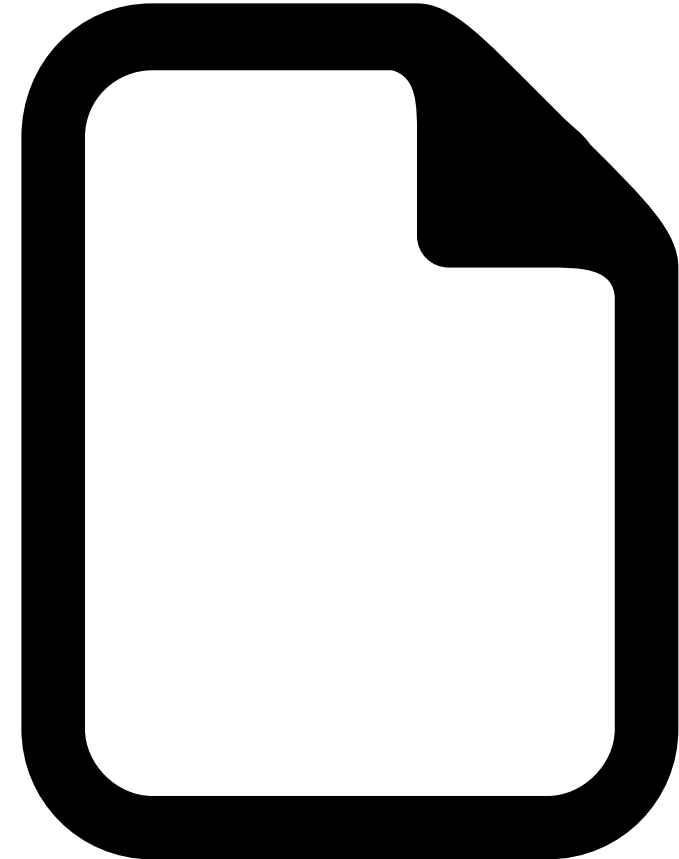
feed, fertilizer, elevator, bins, farmers, seed, mill

Topic 4 Top Words:

mold, castings, sand, aluminum, foundry, pour, cooling

Topic 5 Top Words:

walls, masonry, structural, retaining, waterproofing, dry, basement



What do topics tell us about a document?

Example topics:

Topic 1 Top Words:

farm, crops, tractors, harvesting, acres, plant, grow

Topic 2 Top Words:

drug, testing, checks, required, employment, mvr, physical

Topic 3 Top Words:

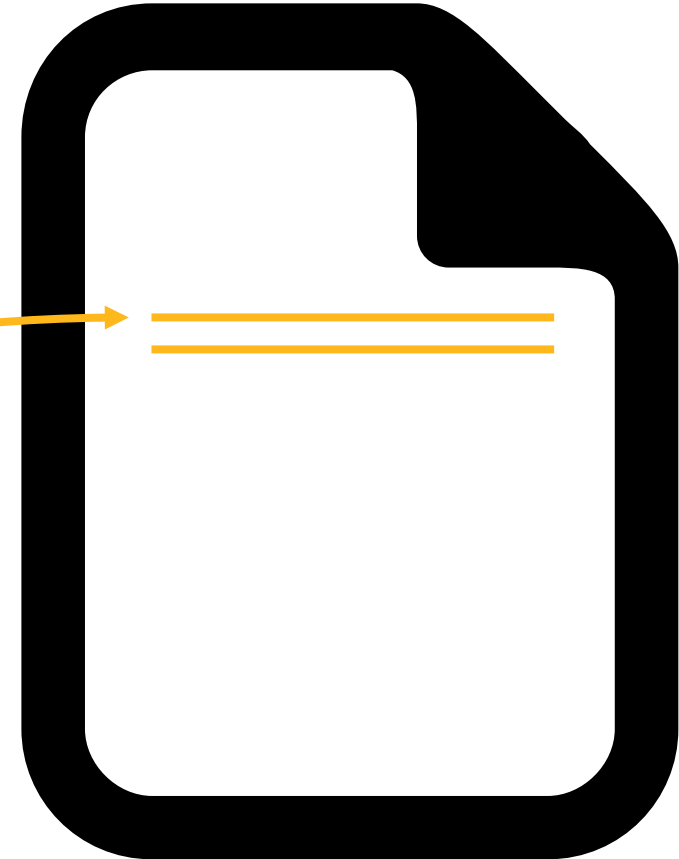
feed, fertilizer, elevator, bins, farmers, seed, mill

Topic 4 Top Words:

mold, castings, sand, aluminum, foundry, pour, cooling

Topic 5 Top Words:

walls, masonry, structural, retaining, waterproofing, dry, basement



What do topics tell us about a document?

Example topics:

Topic 1 Top Words:

farm, crops, tractors, harvesting, acres, plant, grow

Topic 2 Top Words:

drug, testing, checks, required, employment, mvr, physical

Topic 3 Top Words:

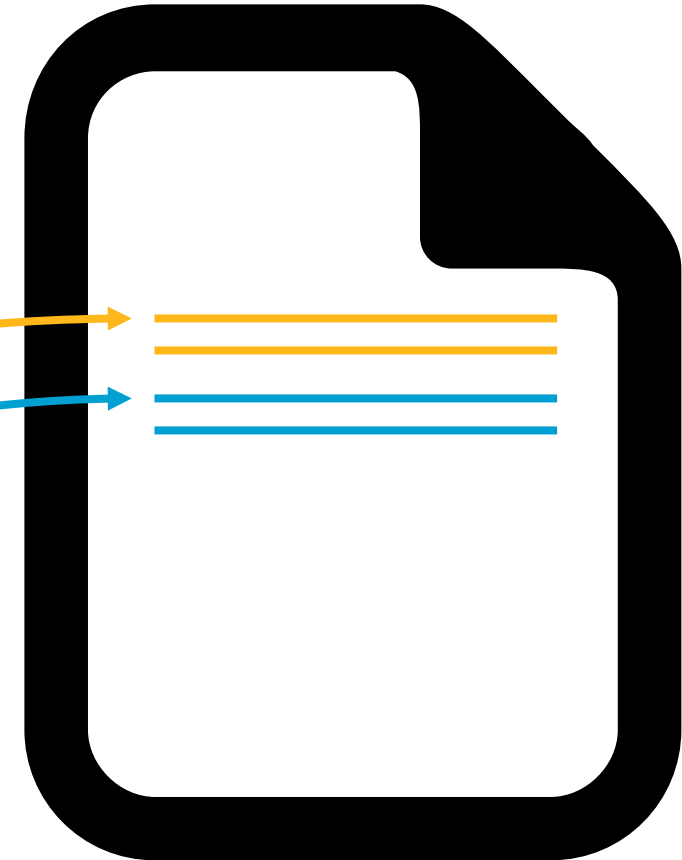
feed, fertilizer, elevator, bins, farmers, seed, mill

Topic 4 Top Words:

mold, castings, sand, aluminum, foundry, pour, cooling

Topic 5 Top Words:

walls, masonry, structural, retaining, waterproofing, dry, basement



What do topics tell us about a document?

Example topics:

Topic 1 Top Words:

farm, crops, tractors, harvesting, acres, plant, grow

Topic 2 Top Words:

drug, testing, checks, required, employment, mvr, physical

Topic 3 Top Words:

feed, fertilizer, elevator, bins, farmers, seed, mill

Topic 4 Top Words:

mold, castings, sand, aluminum, foundry, pour, cooling

Topic 5 Top Words:

walls, masonry, structural, retaining, waterproofing, dry, basement



What do topics tell us about a document?

Example topics:

Topic 1 Top Words:

farm, crops, tractors, harvesting, acres, plant, grow

Topic 2 Top Words:

drug, testing, checks, required, employment, mvr, physical

Topic 3 Top Words:

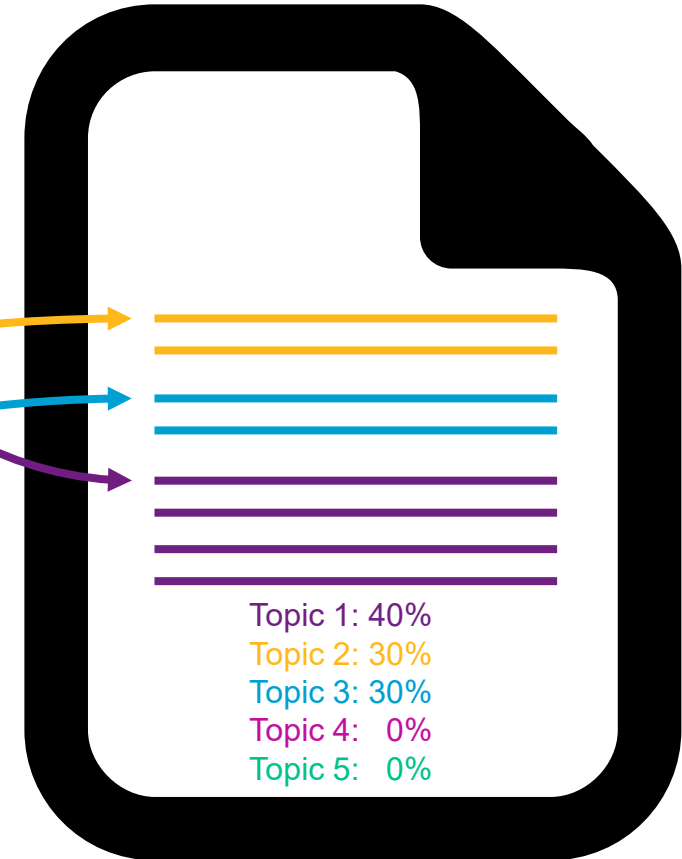
feed, fertilizer, elevator, bins, farmers, seed, mill

Topic 4 Top Words:

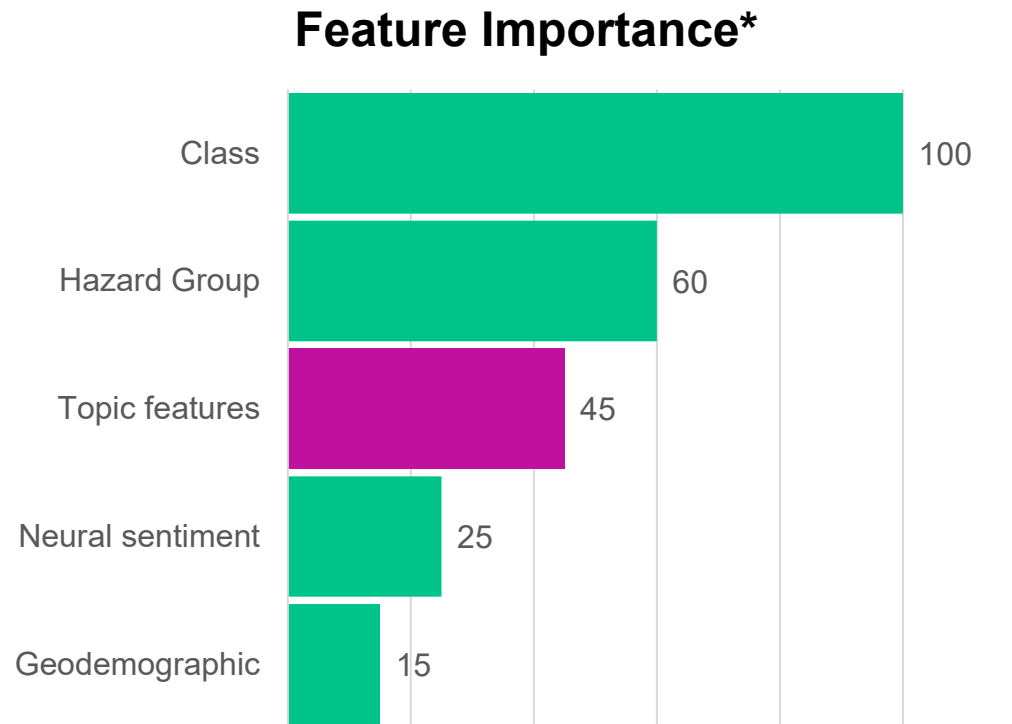
mold, castings, sand, aluminum, foundry, pour, cooling

Topic 5 Top Words:

walls, masonry, structural, retaining, waterproofing, dry, basement



Topics are powerful predictors



* - approximation across multiple models, normalized by Class

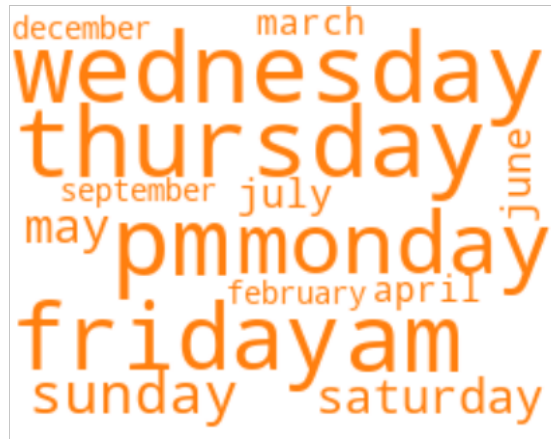
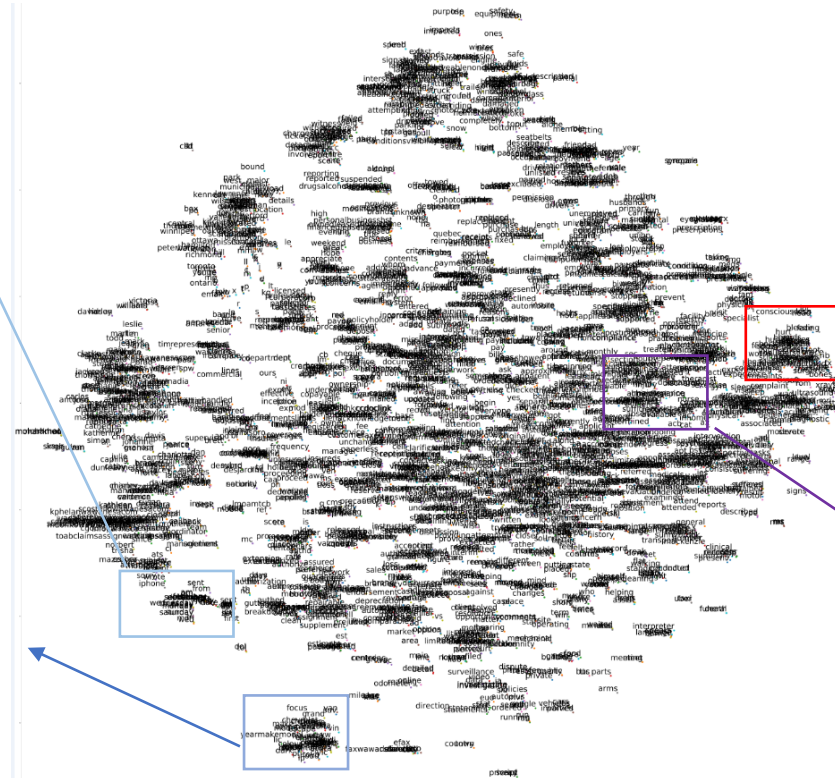
Sample topics

Topic 1	Topic 2	Topic 3
wear	safety	construction
required	training	residential
ppe	safety_program	carpentry
glasses	safety_meetings	framing
safety	documented	remodeling
hats	osha	plumbing
gloves	formal_safety	renovation
safety_glasses	safety_training	siding
boots	written_safety	hvac
hard_hats	certified	subcontract

Case Study – Independent Medical Exams

- An independent medical examination (IME) helps us to:
 - Determine the cause of injury associated with the incident
 - Evaluate the claimant condition and medical treatment
 - Mitigate risk of injury deterioration
- IMEs are paid by the insurance company **in addition to** the medical and rehabilitation costs
 - We want to better understand our current spend
 - We want to optimize IME spend and order only when it is necessary to minimize premium impact for all policyholders

Embeddings - Word to Numbers

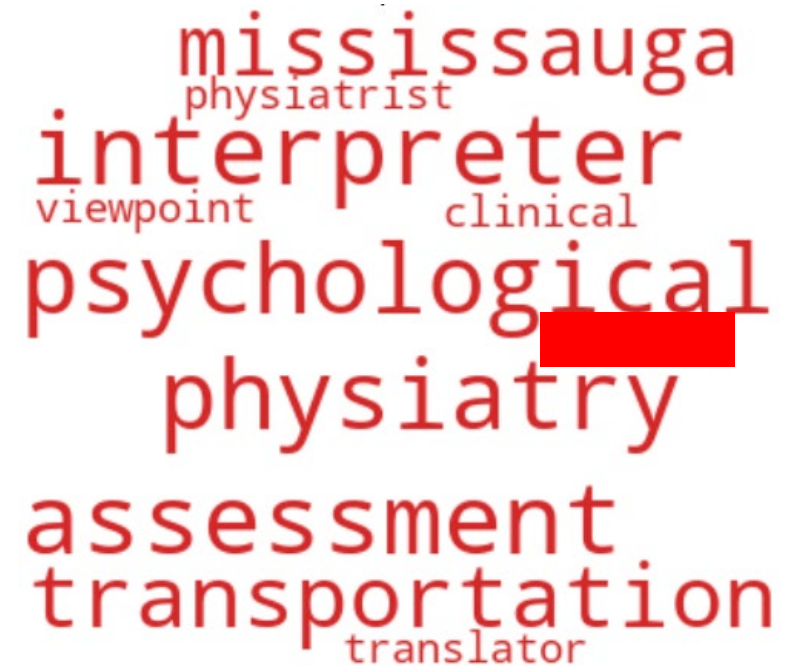


IME – Text Analytics: Classifying Text

- Low probability of IME
- High probability of IME



A word cloud in shades of blue and teal. The words are arranged in a roughly rectangular shape. The most prominent words are 'injuries', 'advised', 'contact', 'unchanged', 'treatment', 'back', 'invoice', 'package', 'transfer', 'insurance', 'plaintiff', and 'police'.



A word cloud in shades of red. The words are arranged in a roughly rectangular shape. The most prominent words are 'mississauga', 'physiatrist', 'interpreter', 'psychological', 'physiatry', 'assessment', 'transportation', 'translator', 'viewpoint', and 'clinical'. There is a red rectangular redaction mark over the word 'psychological'.

Case Study – Who's at Fault?

- In a car accident: who is responsible?
 - We have a structured data column to record fault rating
 - manual entry
 - low quality
 - Can text analytics do better?
- We have subrogation models to predict:
 - Who's insurer should pay for the damages?

At Fault Rating Discrepancy

ClaimCenter - **structured** fields:

General

Fault

Insured not at fault

Insured's Liability %

0

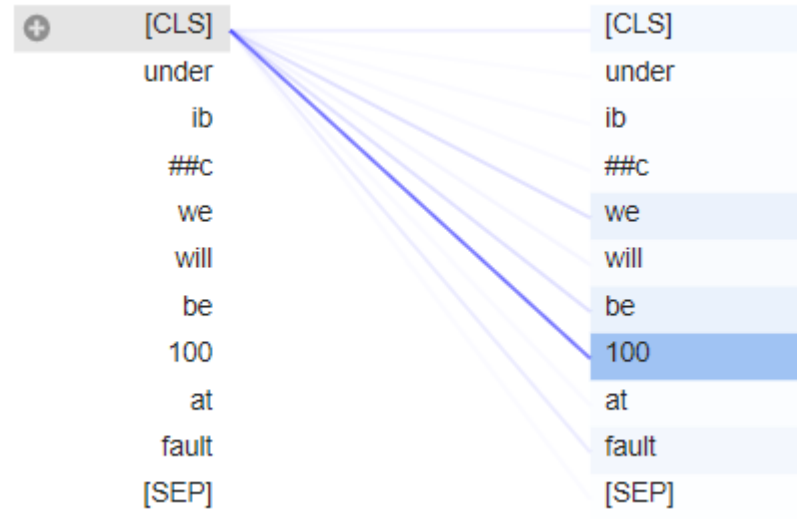
ClaimCenter – **unstructured** notes:

02:20 PM

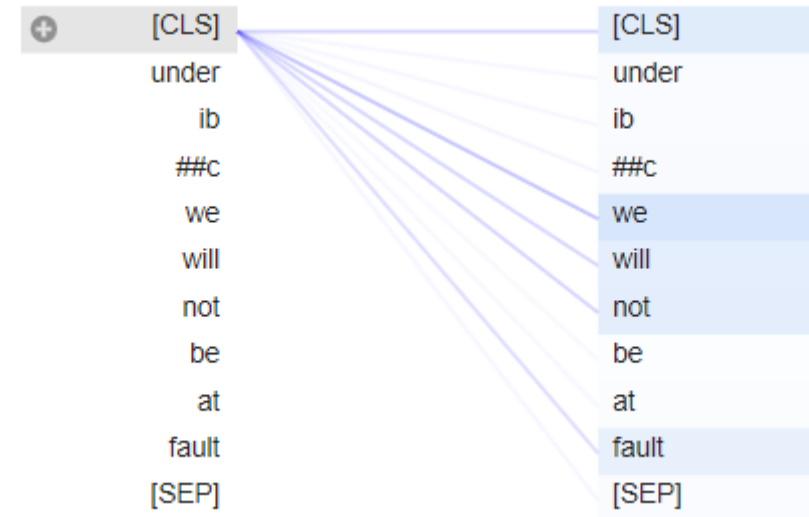
Reviewed claim and the statements from both parties and the estimate advised would appear hit dead center front end therefore TP would of been directly infront of our insured therefore i have gave waive and send for 100% with our insured being held liable TPA advised there insured's vehicle is a TL

BERT– Attention Patterns

“under IBC we will be 100 at fault”



“under IBC we will not be at fault”



InterBERTability

Generate a recommended correction that looks like this:

Claim Number	Insured at Fault	Recommended Insured at Fault	Recommendation Explanation
	0%	94%	unclear of the color of the light the decision was reached in favor with royal with cooperators being 100 at fault left turning vehicle was held at fault with no witnesses to the color of the light ccdoclink 23211544 apd emailed insured going over liability decision and how we will be held 100 liable for the loss included auto liability letter in the email

Captum library (<https://github.com/pytorch/captum>).

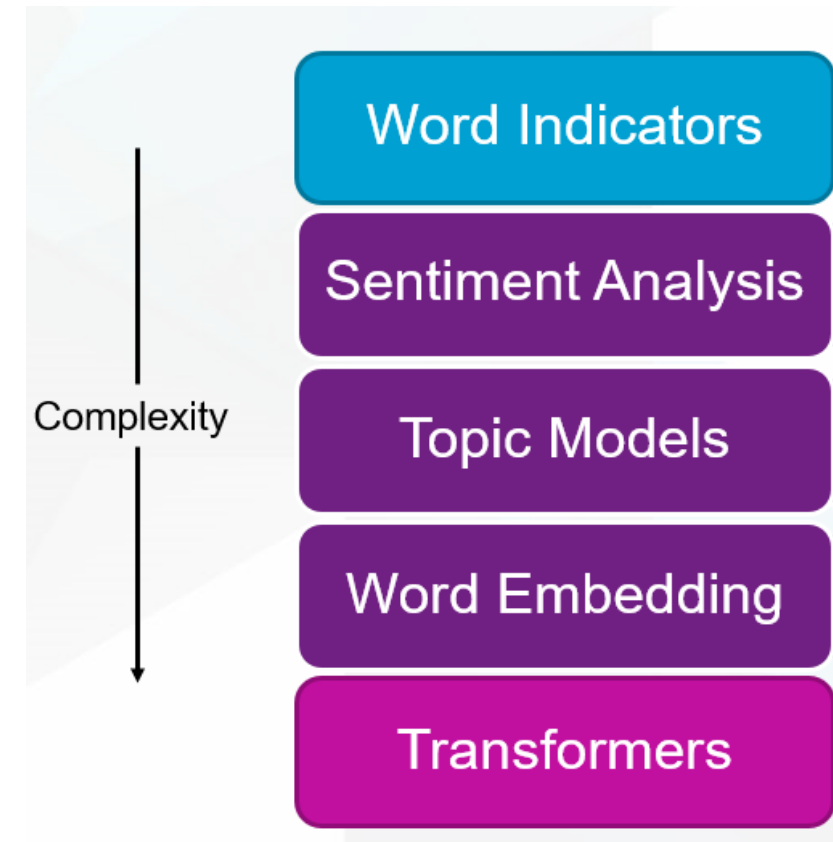
Transformers Library - Huggingface

- Deep interoperability with TensorFlow and PyTorch
- Over 32+ pretrained models in 100+ languages
 - BERT
 - GPT-2
 - RoBERTa
 - XLM
 - DistilBert
 - XLNet
 - CTRL
 - ...



Conclusions

- Variety of techniques
 - Basic to very complex
 - Quickly growing field
- Value of Natural Language Processing
 - Provide insights
 - Augment structured data
 - Make better predictions



Thank you



Willis Towers Watson 