InsurTech in Action

Zhiyu (Frank) Quan

University of Nebraska–Lincoln

2023-10-04

What is InsurTech?

- InsurTech refers to the **application of emerging technology** across the entire **insurance value chain** in order to address existing problems and uncover new opportunities.

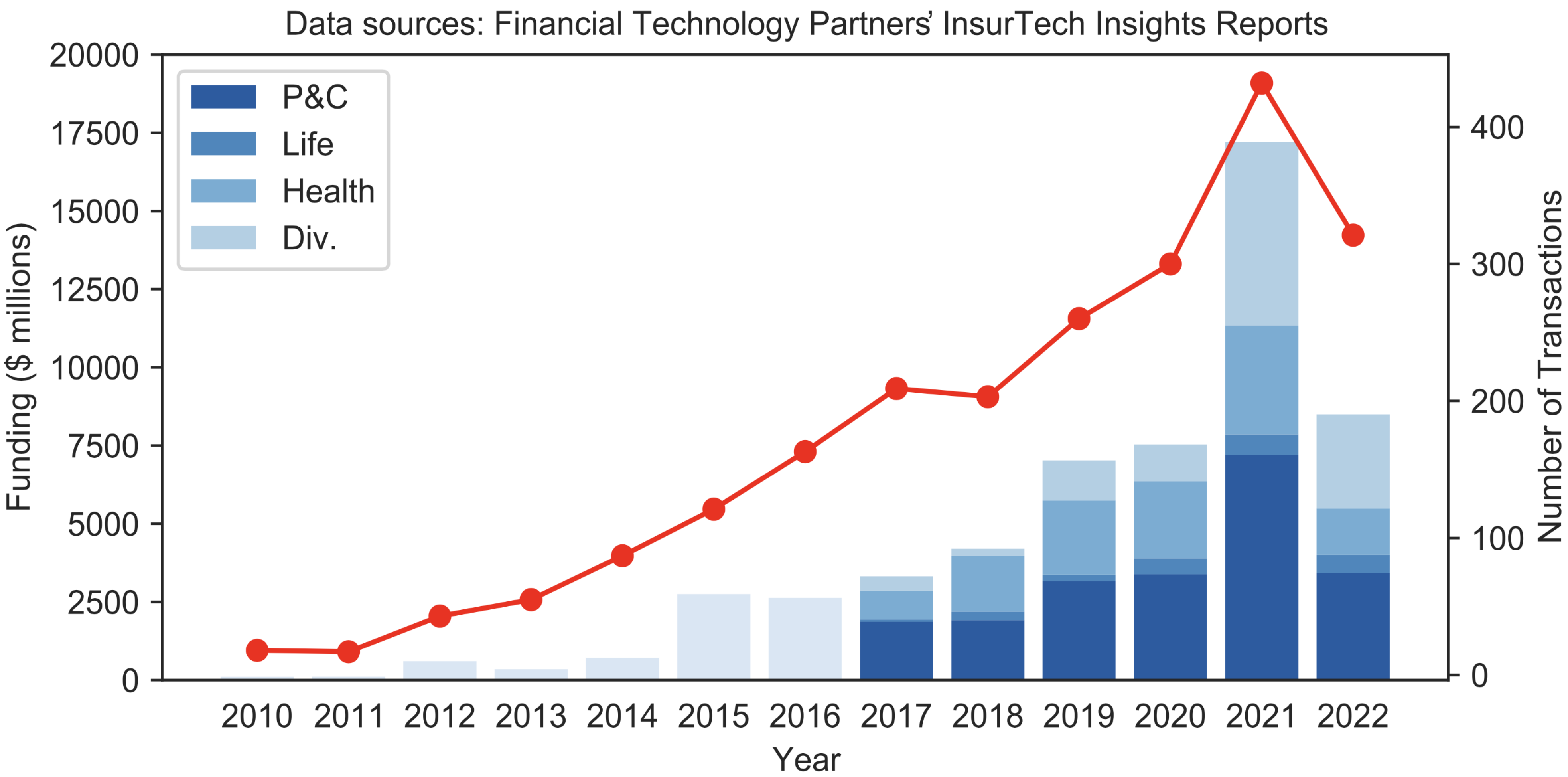
- InsurTech delivers **user-oriented** or **data-driven** solutions to the insurance industry including automation of business processes, the development of innovative products, and the exploitation of data for underwriting, risk assessment and claim handling.

Source of image: https://gomedici.com/taking-pulse-of-insurtech-insurance-india

InsurTech Examples

- **Mobile devices with apps** automate business processes such as reporting claims, purchasing insurance products, and customer service.
- Health insurance company use **wearable techology** such as smart bracelets to track health measurements and reward healthy behaviour for customers seeking a healthier lifestyle.
- Pay-As-You-Drive auto insurance plans or Usage-based insurance (UBI) use **telematics technology** to allow drivers to have premiums tailored to their driving behaviors.
- Life insurance companies use **facial analytics technology** to produce a life insurance quote by analyzing a selfie photo and estimating the relevant data of the insured (e.g., BMI).

InsurTech Financing for 2011-2022



Improving Business Insurance Loss Models by Leveraging InsurTech Innovation

Joint work with Changyue Hu, Panyi Dong, Emiliano Valdez

Data and Risk Analytics with InsurTech

Technology innovation	Insurance type	In-house (traditional) factors	InsurTech factors
Wearables 	Life & health	• Age, gender, marital status • Pre-existing medical condition • Family history • Body mass index • Tobacco use	• Blood pressure • Heart rate • Glucose level • Frequency of exercise • Sleep pattern
Telematics 	Auto	• Age, gender, marital status • Driving history • Credit rating • Type of car • Business or pleasure • How much you drive	• Brakes • Acceleration • Rotation and turns • Location, weather condition • Distance traveled • Driving attentiveness
Smart Homes 	Home	• Alarm and security system • Age, home structure • Home square footage • Type of roof • Fire safety and protection	• Window and door sensors • Smart thermostats • Smart locks • Smoke detector • Water and leak detection

Objective

- The goal is to build a predictive loss model for XYZ Insurer's BOP line of business by leveraging innovative data sources from Carpe Data, an InsurTech company.

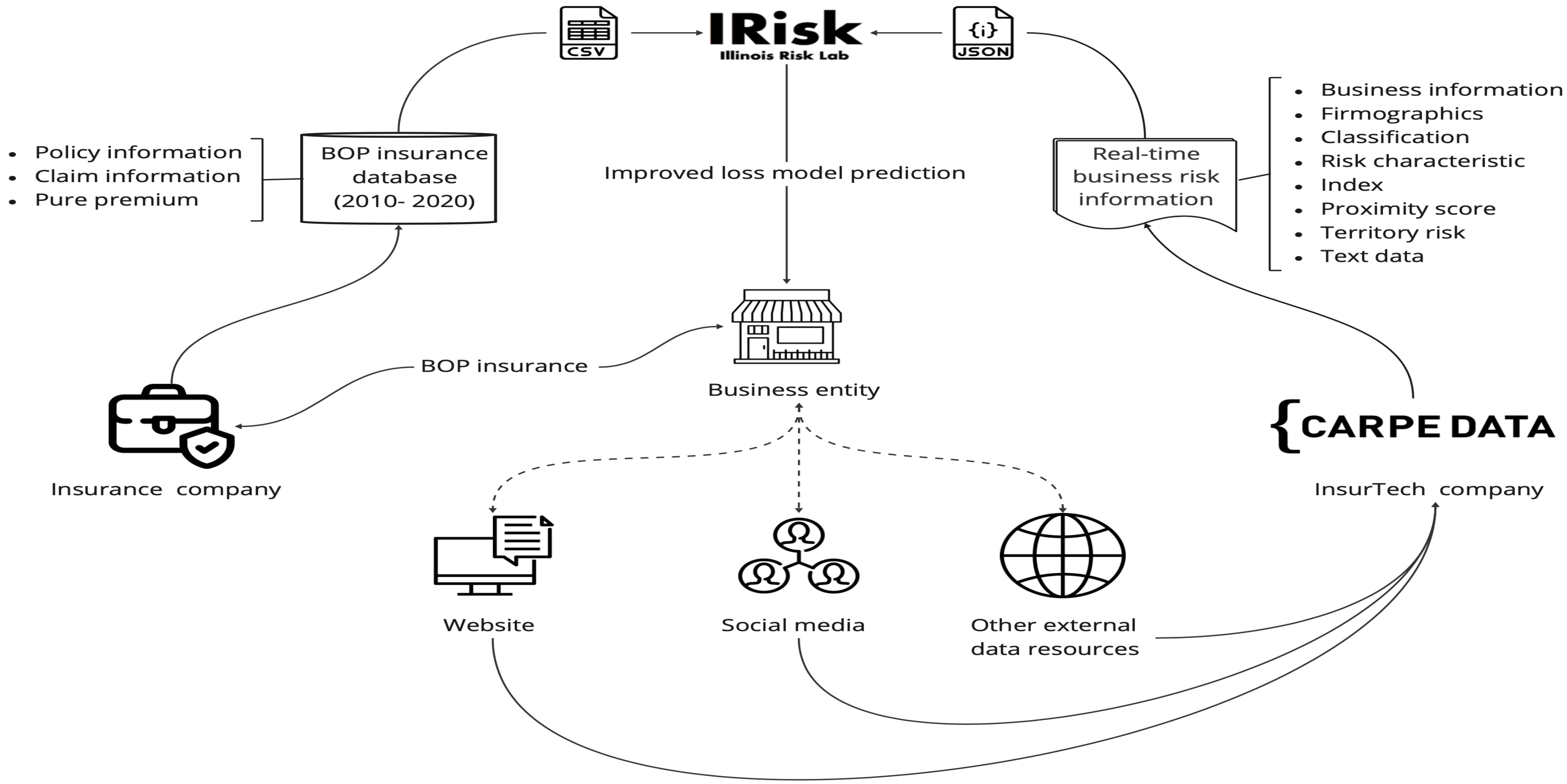
Business Owner's Policy (BOP)
Loss Data (2010 - 2020)

Data contains policy information and rating factors used in-house about each insured business.

Supplemental data source
for business-related features

Features describing insured businesses created by InsurTech Innovation

Industry & University Collaboration: IRisk Lab



Business Owner’s Policy (BOP) Insurance

- Business Owner's Policy (BOP) is a commercial/business insurance policy intended to protect **small** or **medium-sized** business owners against potential risks.
- It is available and applicable to **variety of industries**.
- It combines a **various insurance coverages**, typically including general liability insurance, commercial property insurance, and business interruption insurance.
- The **complexity of underwriting and claims**, combined with the **low volume and bespoke nature of transactions**, pose **obstacles** to commercial insurance embracing InsurTech.

Insurance Data from XYZ Insurer

- XYZ's **BOP historical loss experience from 2010 to 2020**.
- More than **1,200,000** data entries.
- **Policy information**:
  - Policy Year
  - Earned Exposure
  - Coverage Limit
  - Exposure Base: *LOI, Annual Gross Sales, Annual Payroll*
  - Risk Type: *Apartment, Condo/Office, Contractors, Convenience, Distributor, Fast Food, Motel, Office, Other, Restaurant, Retail, Self-Service*
  - Coverage Type: *Building (BG), Business Personal Property (BP), Liability (LIAB)*
- **Loss experience and in-house predictive model**
  - Observed Loss Cost
  - Insurance Company's In-house Model Loss Cost

InsurTech Data from Carpe Data / InsurTech Innovations

Carpe Data provided us with **real-time**, **dynamic** information from emerging **public data** sources shed lights on numerous facets of a business: operations, products, services, physical plant, etc.

- **Business Information**: General operation information about a business.
  - *is_home_business, founded_year, opening hours, description*
- **Firmographics**: Characteristics to segment prospect business.
  - *business Size, company type, revenue range*
- **Risk Characteristics**: Various risk attributes of a business.
  - *commercial cooking equipment, raw seafood and alcohol (for a Japanese restaurant)*

InsurTech Data from Carpe Data / InsurTech Innovations

- **Classification**: Categorization of a business.
    - *category, segment, NAICS code*
- **Reviews**: Available reviews for a business.
    - *reivew content, # likes, star rating, response from owner*
- **Webpage**: Details on a business's webpage.
    - *content, title, url*
- **Group**: Features engineered from collected information.
    - *group 1 - group 13*

InsurTech Data from Carpe Data / InsurTech Innovations

- **Next-generation scores & indexes** A suite of indexes on a 1 – 5 scale targeting dimensions of risk that can be tuned by segment and location.

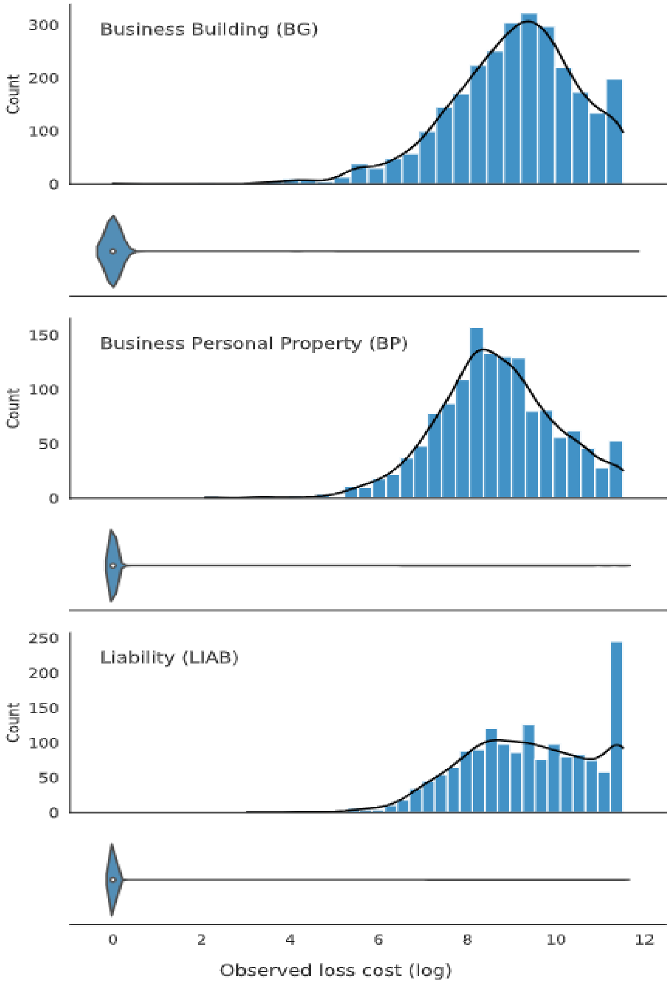
Scores:

- *Negative Keywords*
- *Proximity Combustibles*
- *Proximity Entertainment*
- *Proximity Traffic Mode*

Indexs:

- *Customer Rating*
- *Visibility*
- *Reputation*
- *Health & Sanitation*
- *Maintenance & Condition*

Distribution of Observed Loss Cost



Observed loss cost:

- *Imbalance*
- *Heavy tail*
- *Differ across coverages*

Modeling - Light Gradient Boosting Machine (LGBM)

- A **gradient boosting** framework that uses **tree-based** learning algorithms.
- **Advantages** of Light GBM
    - Faster training speed and higher efficiency.
    - Lower memory usage.
    - Support parallel, distributed, and GPU learning.
    - Capable of handling large-scale data.

Hyper Parameter Tuning - Bayesian Optimization

- **Bayesian optimization** can effectively narrow the hyperparameter space.
  - Use previous evaluation results to choose the next optimal hyperparameters to evaluate.
- **Distributed learning** by **optuna**
  - Multiple batch jobs on the same model with different sets of parameters.
  - Multiple parameter sets are being trained simultaneously.

OPTUNA

Double Lift Charts

- Model lift refers to the ability to **differentiate between low and high lost policyholders** and can be used to measure a model's economic worth.

- Double lift charts are commonly used to measure the model lift and **compare the predictiveness between two different models**.

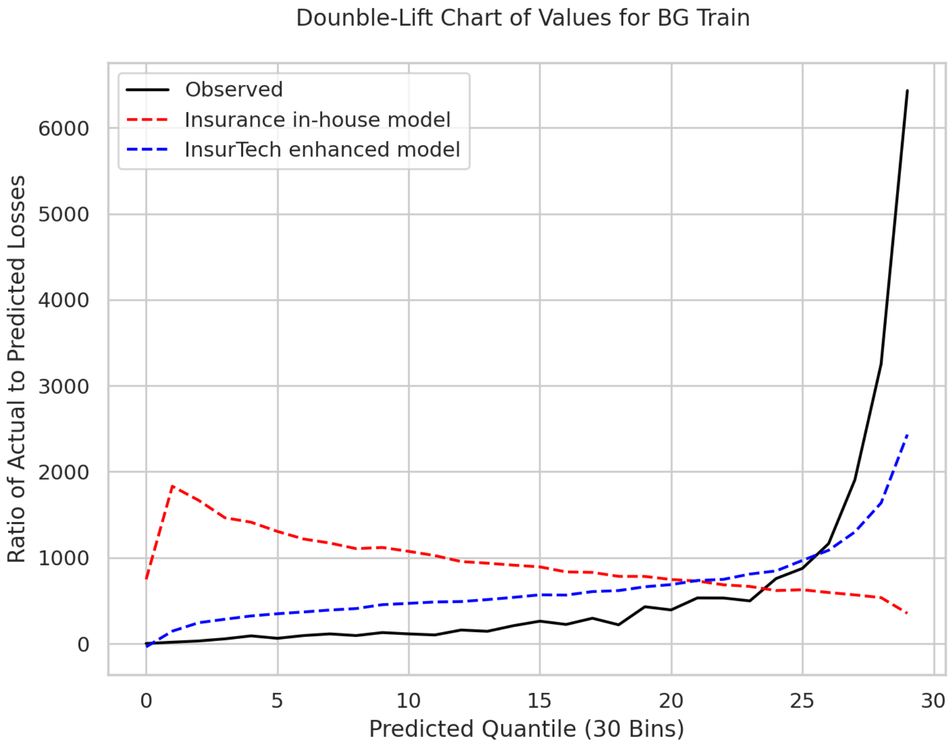
- Double lift charts are created as follows:

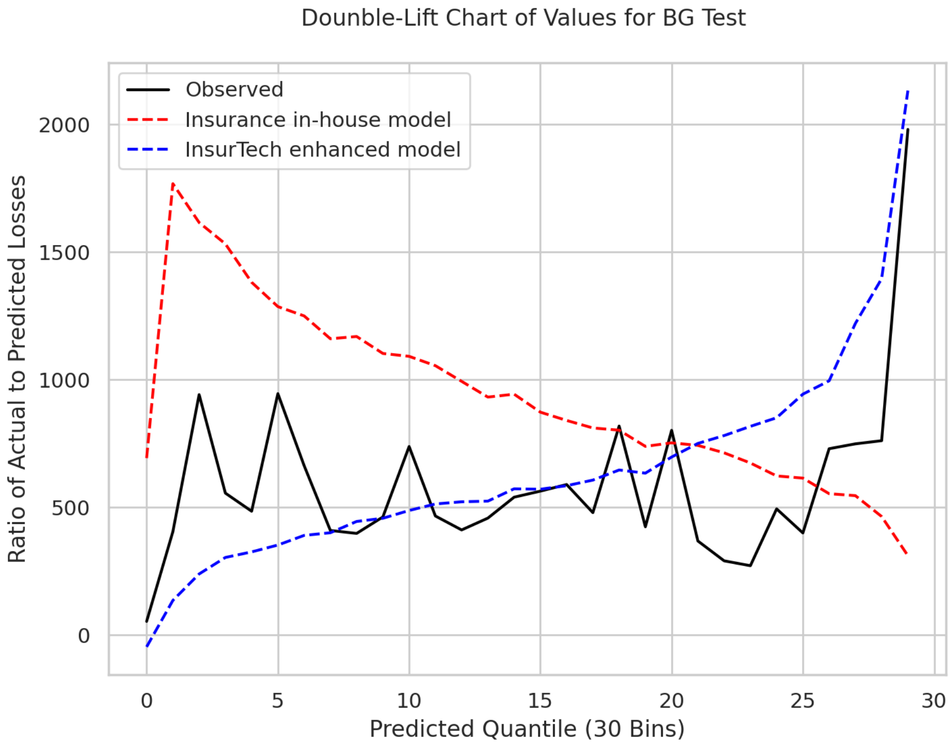
    - Sort data by a ratio of new model prediction (Insurtech-enhanced model prediction) to the current premium (insurance in-house model prediction).
    - Subdivide sorted data into quantiles with equal exposure (we use 30 quantiles).
    - For each quantile, calculate the average observed loss, the average current premium (insurance in-house model prediction), and the average new model predicted loss (Insurtech-enhanced model prediction).

- The model that gives **better predictions** is the one whose predicted loss line is **closer to the observed** one.

Double Lift Charts

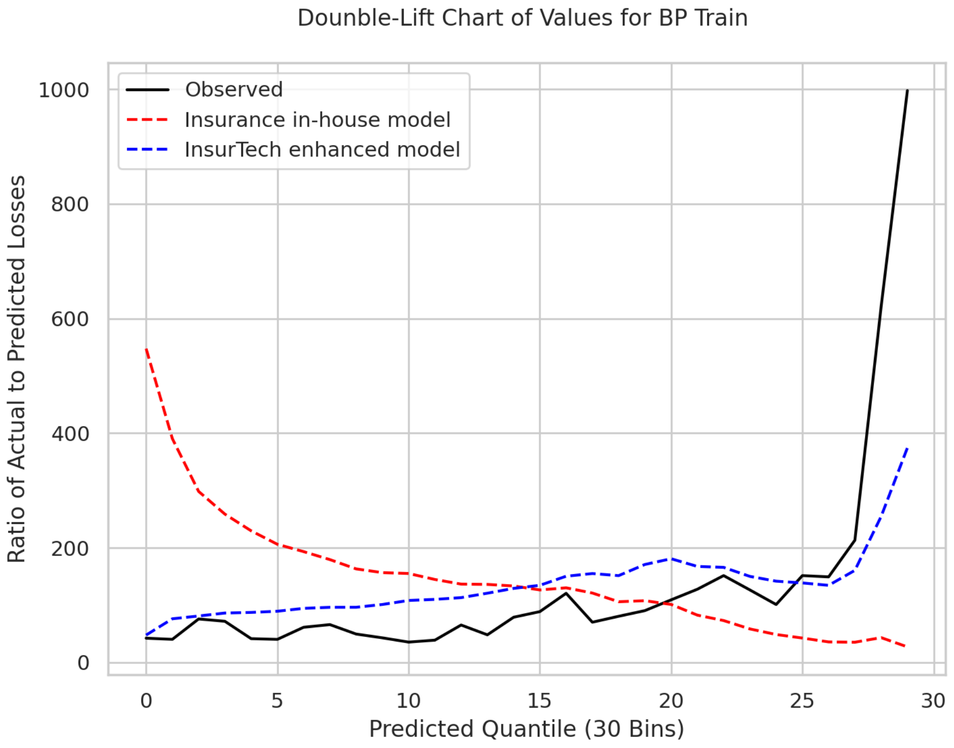
Building LGBM MAE Model

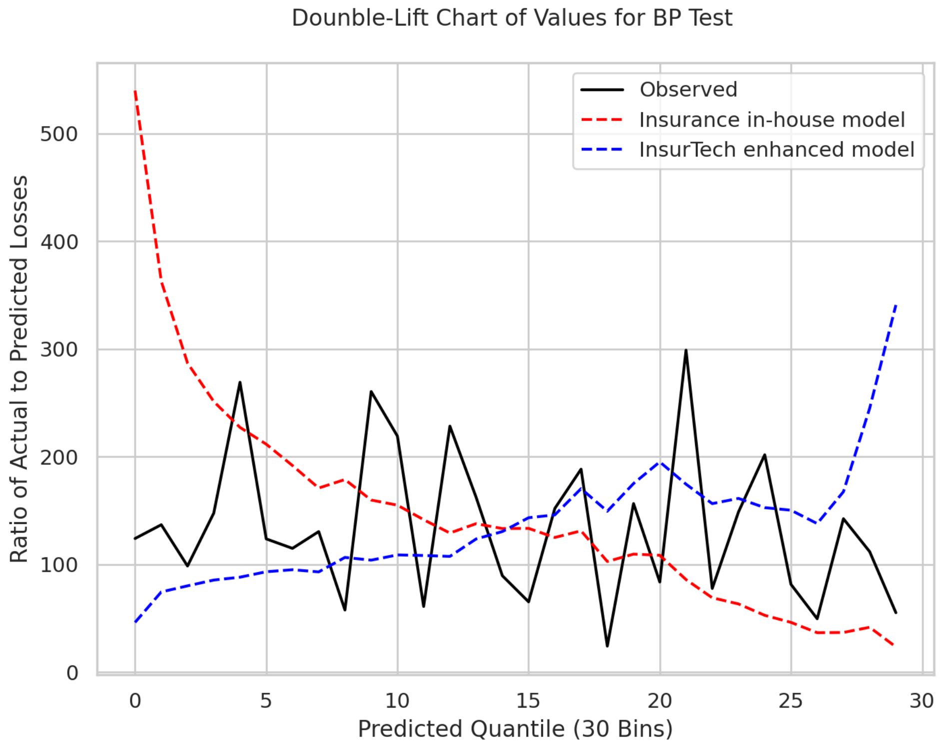




Double Lift Charts

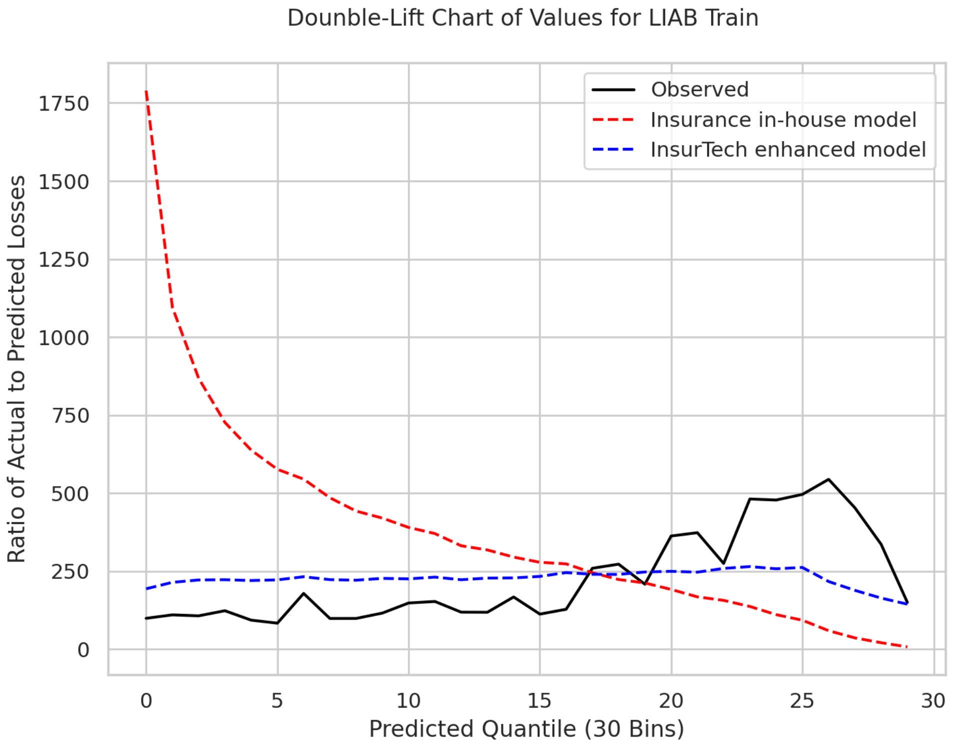
Business Personal Property LGBM MAE Model

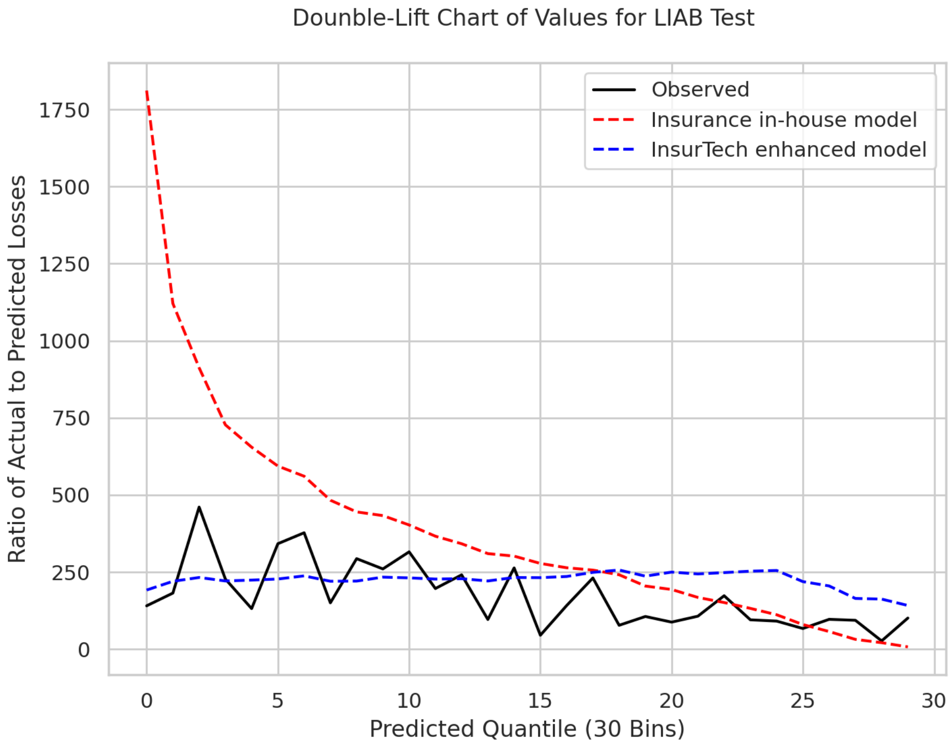




Double Lift Charts

Liability LGBM MAE Model

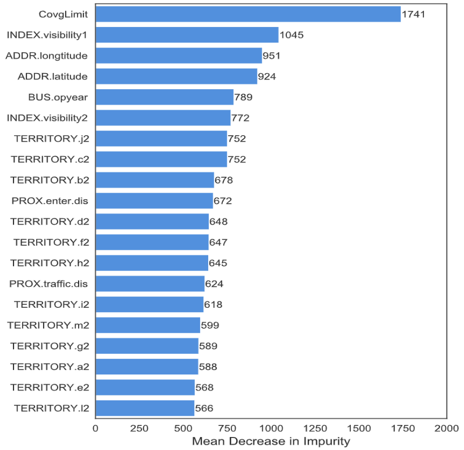




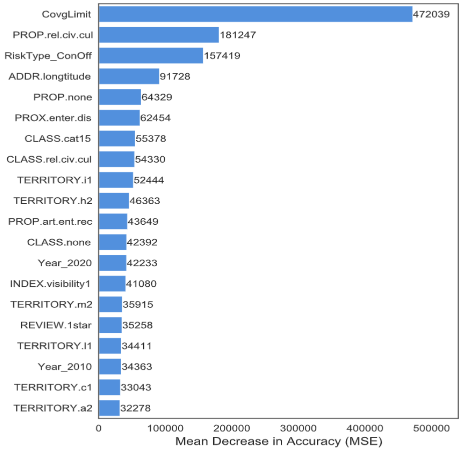
Model Performance based on Validation Measures

Coverage	Dataset	Model	Gini	PE	RMSE	MAE
BG	train	Insurance in-house model	0.29	-0.40	5761.94	1526.47
		Tweedie GLM + elastic net	0.44	-0.04	5660.01	1286.31
		LightGBM	0.84	0.00	5364.05	1198.07
	test	Insurance in-house model	0.32	-0.54	5328.02	1461.92
		Tweedie GLM + elastic net	0.32	-0.16	5284.90	1238.94
		LightGBM	**0.37**	**-0.08**	**5198.57**	**1181.47**
BP	train	Insurance in-house model	0.59	-0.07	2498.13	277.37
		Tweedie GLM + elastic net	0.68	0.00	2450.82	262.64
		LightGBM	0.78	0.00	2409.88	259.11
	test	Insurance in-house model	0.58	-0.11	2350.80	270.75
		Tweedie GLM + elastic net	0.36	**-0.04**	2375.10	**262.31**
		LightGBM	**0.59**	-0.06	**2348.93**	262.78
LIAB	train	Insurance in-house model	0.57	-0.67	3937.22	586.88
		Tweedie GLM + elastic net	0.63	-0.04	3920.13	449.25
		LightGBM	0.78	0.00	3853.67	435.65
	test	Insurance in-house model	0.54	-1.15	3347.60	547.02
		Tweedie GLM + elastic net	0.47	-0.33	3340.86	408.15
		LightGBM	**0.56**	**-0.26**	**3305.85**	**394.56**

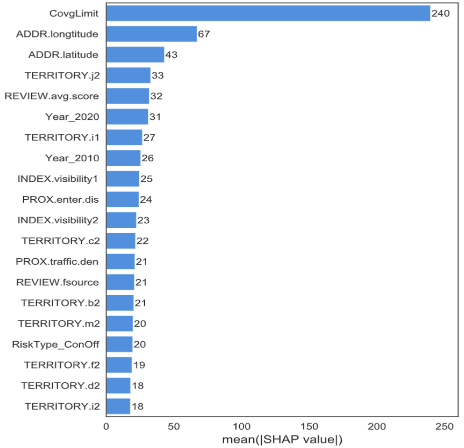
Feature Importance - Top 20 Features - Building



(a) Mean Decrease in Impurity - Top 20 features



(b) Mean Decrease in Accuracy- Top 20 features



(c) SHAP feature importance - Top 20 features

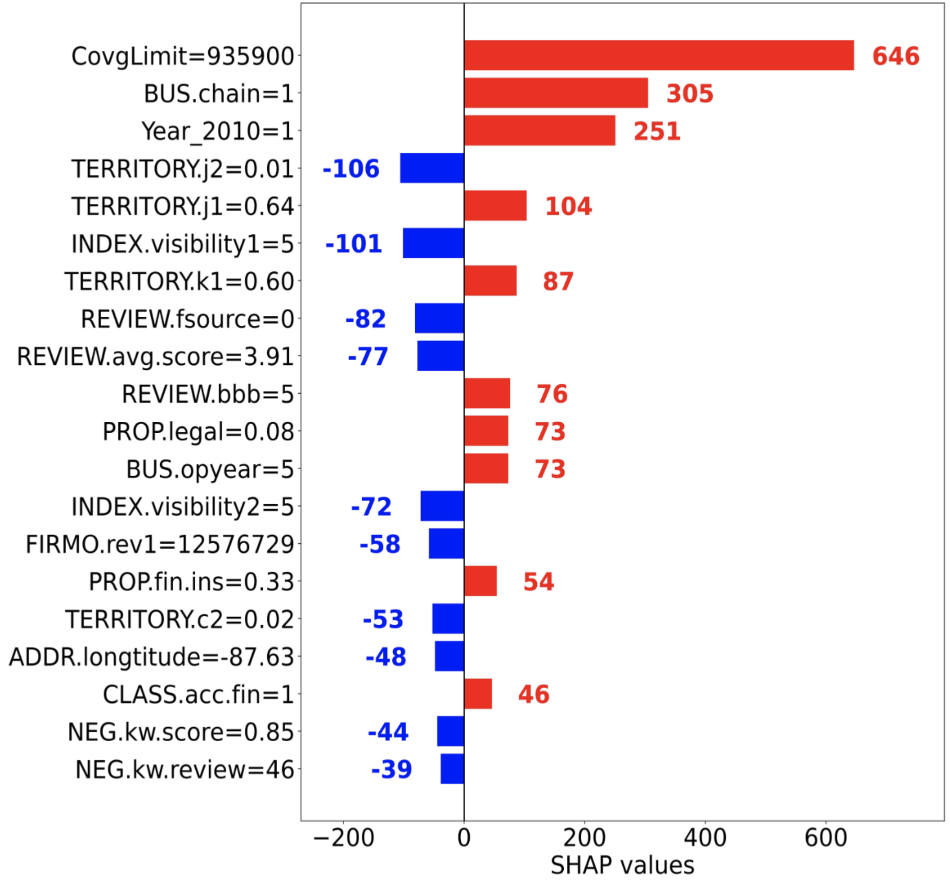
Feature importance:

- Three distinct methods for assessing feature importance.
- With the exception of coverage information, all other significant variables originate from InsurTech.

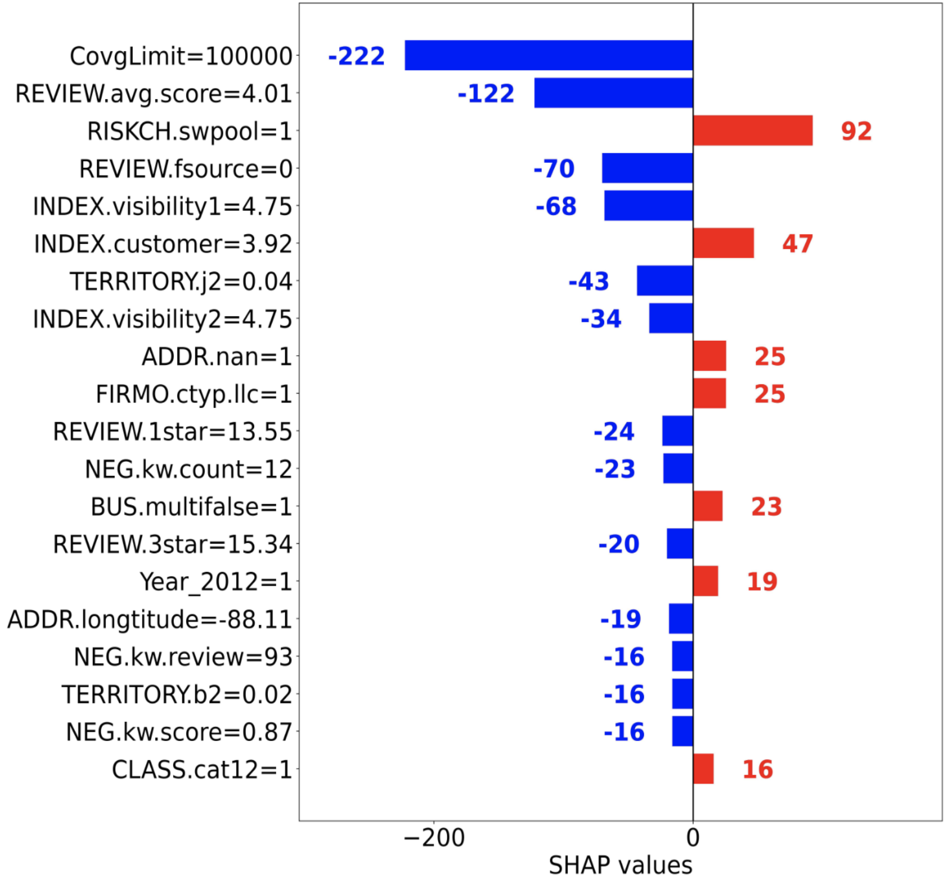
Illustrative Individual Cases

- To further examine how the InsurTech risk factors **affect** the loss model, we extracted and analyzed four real businesses from a **microscopic** point of view.
- Four businesses analyzed are described as follows:
  - (a) a business with **a positive** claim from the **training** dataset;
  - (b) a business with no claim from the training dataset;
  - (c) a business with a positive claim from the test dataset;
  - (d) a business with **no claim** from the **test** dataset.

A-Land Trust Company B-Rental Apartment

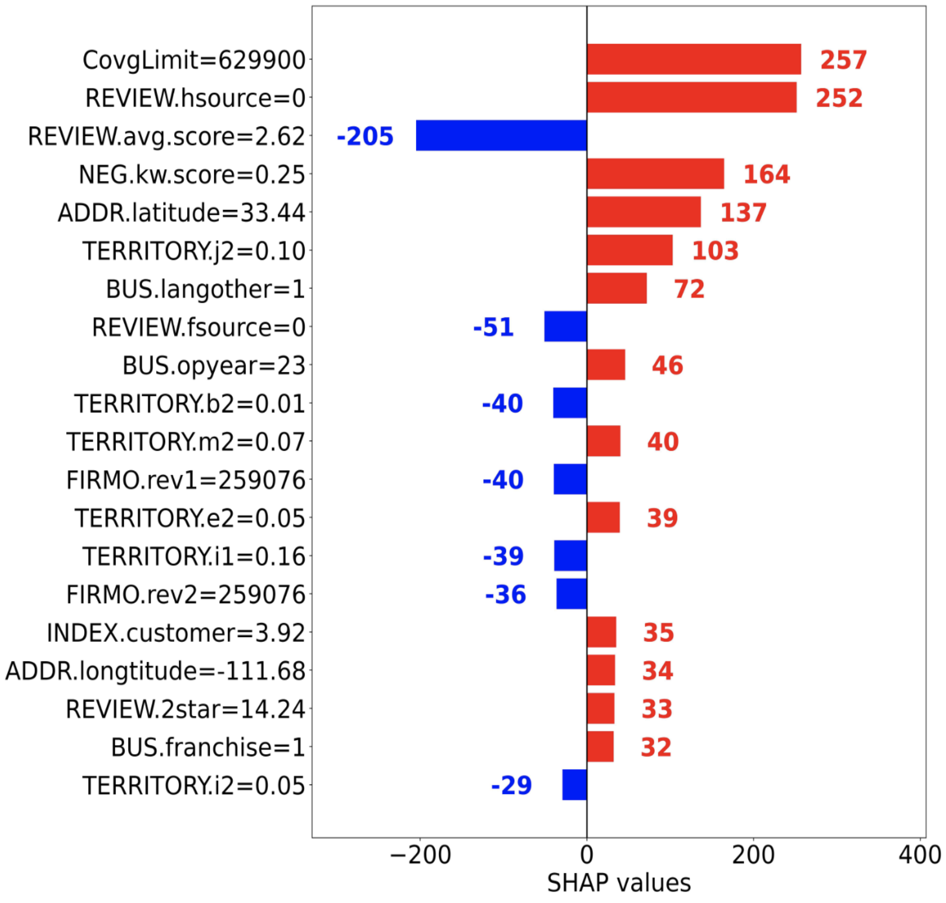


(a) Top 20 influential features of Business A

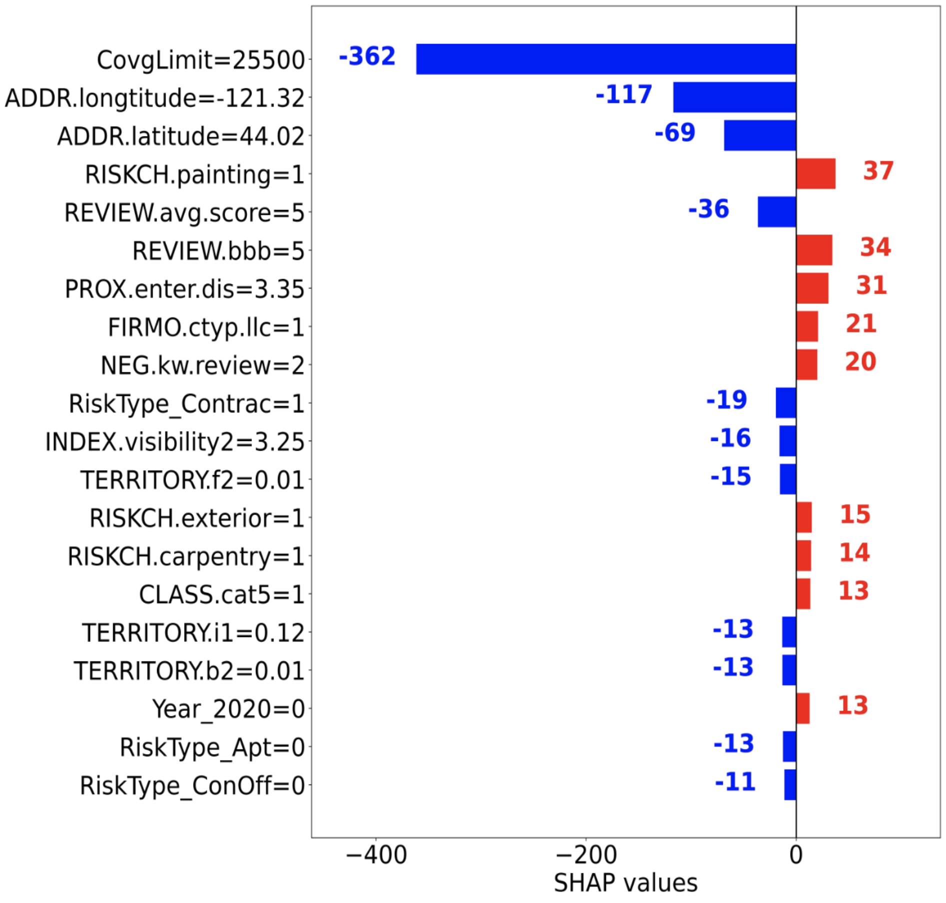


(b) Top 20 influential features of Business B

C-Licensed Medical Clinic D-Contractors



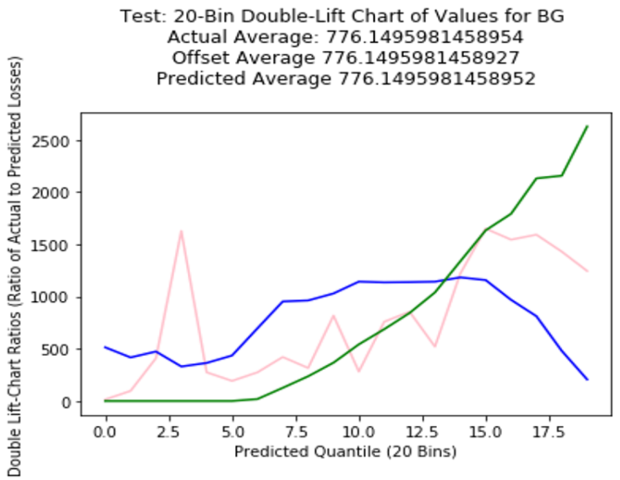
(c) Top 20 influential features of Business C

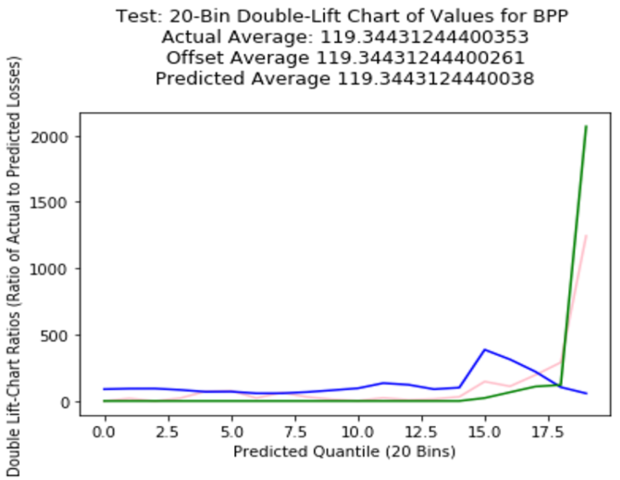


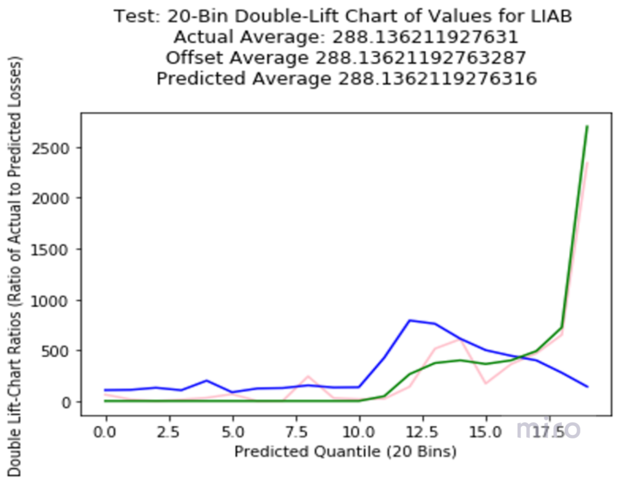
(d) Top 20 influential features of Business D

Implementation Perspective - Residual Modeling

- **Ratio Residual** = Observed Loss Cost / Insurance Company's In-house Model Loss Cost
- Residual modeling has the advantage of improving the predictions **without creating a completely new model**.
- New predictions, however, will be **heavily influenced by old predictions**.







Future Steps

- Insurance tailored feature engineering
  - NLP techniques on text data from social media
  - Spatial and temporal modeling on foot traffic data
- Other modeling approaches
  - Hu, C., Quan, Z., & Chong, W. F. (2022). Imbalanced learning for insurance using modified loss functions in tree-based models. Insurance: Mathematics and Economics, 106, 13-32.
  - Quan, Z., Wang, Z., Gan, G., & Valdez, E. (2023). On hybrid tree-based methods for short-term insurance claims. Probability in the Engineering and Informational Sciences, 37(2), 597-620.

Concluding Remarks

- InsurTech helps enhance loss model predictions using their databases, otherwise **inaccessible** by insurers, to gain better insights into the underlying risks.
- This project aims to investigate how much **improvement** can be gained from these resourceful data.
- Our results indicate substantive differences in the loss cost predictions using **real-life** data from an insurer’s portfolio of BOP policies.
- This work is an example of the benefits that can be gained from a successful **industry and university collaboration** through the Illinois IRiskLab.

Selected References

- Carrie, K. and Kiki, W. (2021). Insurtech : A guide for the actuarial community. Technical report, Willis Tower Watson.
- Chester, A., Hoffmann, N., Johansson, S., and Olesen, P. B. (2018). Commercial lines insurtech: A pathway to digital. McKinsey & Company.
- Eling, M. and Lehmann, M. (2018). The impact of digitalization on the insurance value chain and the insurability of risks. The Geneva papers on risk and insurance-issues and practice, 43(3):359–396.
- Hernan, L. M. (2016)And The Winner Is...? How to Pick a Better Model. https://www.casact.org/sites/default/files/presentation/rpm_2016_presentations_pm-lm-4.pdf
- Naylor, M. (2017). Types of insurance. Insurance Transformed, pages 41–45.
- Parodi, P. (2015). Pricing in General Insurance. Chapman and Hall/CRC, 1 edition.
- Steiner, K. and Meng, B. (2019) Predictiveness vs. Interpretability https://www.soa.org/globalassets/assets/library/newsletters/compact/2019/october/2019-compact-iss63-steiner-meng.pdf

Selected References

- Stoeckli, E., Dremel, C., and Uebernickel, F. (2018). Exploring characteris- tics and transformational capabilities of insurtech innovations to understand insurance value creation in a digital world. Electronic markets, 28(3):287–305.
- VanderLinden, S. L., Millie, S. M., Anderson, N., and Chishti, S. (2018). The In- surtech book: The insurance technology handbook for investors, entrepreneurs and fintech visionaries. John Wiley & Sons.
- Xu, X. and Zweifel, P. (2020). A framework for the evaluation of insurtech. Risk Management and Insurance Review, 23(4):305–329.
- Yan, T. C., Schulte, P., and Chuen, D. L. K. (2018). Insurtech and fintech: banking and insurance enablement. Handbook of Blockchain, Digital Finance, and Inclusion, Volume 1, pages 249–281.

NLP-Powered Repository and Search Engine for Academic Papers

A Case Study on Cyber Risk Literature with CyLit

Joint work with Changyue Hu, and Linfeng Zhang

SOA Report

https://www.soa.org/resources/research-reports/2023/cylit-nlp-search/

https://cylit.math.illinois.edu/

Motivation

The need for advanced literature repository

- No centralized repository of cyber risk literature
  - **Limited coverage**, e.g., Web of Science & Scopus, Martin-Martin et al. (2018)
- Lack of contextual awareness tool for finding cyber risk literature
  - **Keyword-based search**, e.g., Google scholar, Beel and Gipp (2009)
- Insufficient integration of the trends in research
  - **Static** nature and **manual review** processes of survey papers, e.g., Eling (2020)

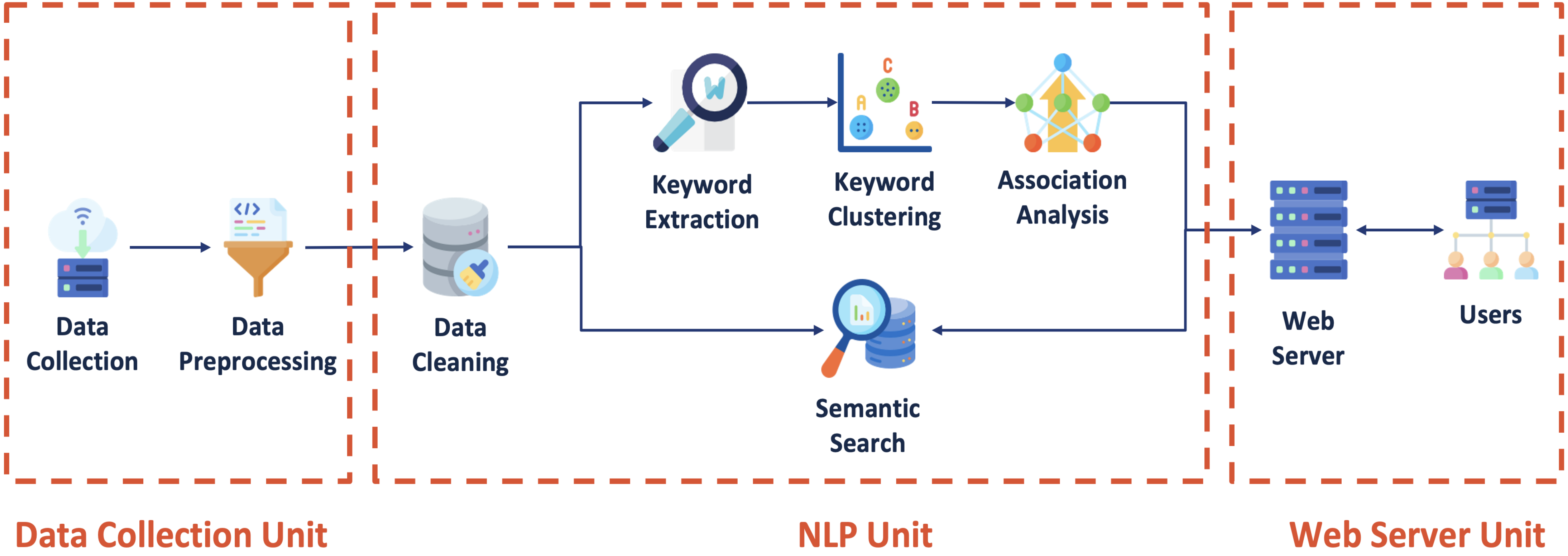
Our Solution and Contribution

Leveraging Natural Language Processing (NLP) techniques to automate literature retrieval, summarization, and classification.

- A comprehensive framework with website interface - living literature database and tailored academic search engines
- Application of NLP techniques to improve efficiency and effectiveness
- Applicability demonstrated in the cyber risk domain

Overview

System Architecture



Potential Upgrades using NLP

University of Nebraska–Lincoln Global Research Rankings of Actuarial Science and Risk Management & Insurance™ released by the College of Business at Nebraska.

- Research trend
- Facilitate collaboration
- Dynamic ranking system: innovation and creativity, interdisciplinary connections, public and practical impact, etc.
- ...

Thank you! Q&A