Enhancing Business Insurance Loss Models through InsurTech Innovations

Emiliano A. Valdez, PhD, FSA
University of Connecticut

Joint work with Z. Quan, C. Hu and P. Dong (U. of Illinois at Urbana-Champaign)

The insurance industry and digital revolution

- Disruptive technologies are evolving, affecting every sector of the economy, including the insurance industry, and creating challenges in their implementation.

- Unlike their other counterparts in the finance sector, the insurance industry has hardly changed for over three centuries, in spite of rapid progress in digital innovation.

- This lack (or slow pace) of innovation in the insurance sector has been attributed to:
  - Industry is very complexed: varied products, how offerings are priced, actuarial pooling of homogeneous risks, transfer of large risks
  - Industry is heavily regulated: rates cannot be excessive but sufficient to remain solvent, rates cannot discriminate against protected classes
  - Profits have historically been stable: why change now?

UCONN.

InsurTech

What is InsurTech?

- the innovative application, adoption, and implementation of existing (evolving) technologies;
- to improve efficiency and drive cost savings in the insurance industry

- It is a departure from existing conventional insurance models of underwriting, risk pooling, pricing, and management.
- It is inspired and derived by "FinTech," a similar innovation that has transformed the banking sector.
- The first such wave of InsurTech happened around 2010.

UCONN
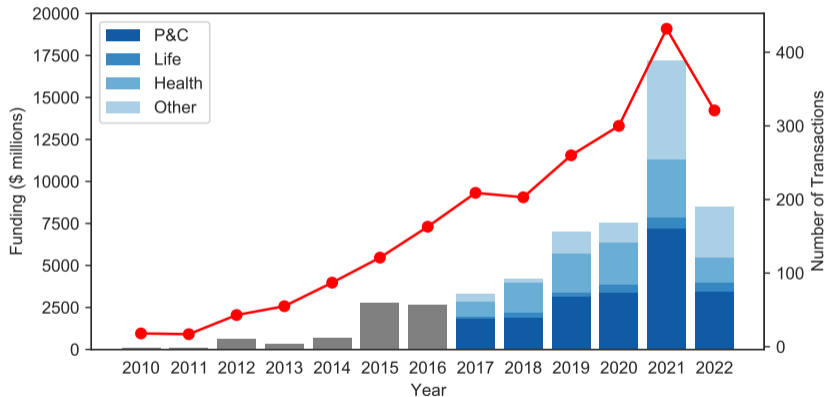
- Total funds invested, $14.4 billion, in 2021 smashed records of amounts invested in 2019 and 2020 combined.

- Today, there are over 1,500 start-up InsurTech deals, with over 500 of them in 2021.

UCONN.

Examples of InsurTech companies

- Oscar https://www.hioscar.com/
  - Company offers health insurance products to individuals, families, and small businesses.
  - "Customer-centric" approach to health care by providing suitable health plans
  - Company uses technology and data to build proactive models to optimize patient-doctor relationship.

- Hippo Insurance https://www.hippo.com/
  - Company offers home insurance using modern technology.
  - Company aims to extend a smarter coverage, much simpler procedure, and innovative ways for taking care of people's home.

- Carpe Data https://www.carpe.io/
  - Company uses AI technology to harness the power of emerging, alternative data for insurers.
  - Insurers can use their database to gain further insights into their underlying risks to enhance underwriting, risks assessments, and claims management.

UCONN

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

UCONN

Project overview

The goal is to build a predictive loss model for XYZ Insurer's BOP line of business by leveraging innovative data source 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

UCONN

Business Owner's Policy (BOP) Insurance

- BOP is insurance for business owners:
  - extends safeguards for your buildings, machineries, equipments, and product inventory (commercial property);
  - provides protection from liability claims and lawsuits (general liability); and
  - covers financial losses arising from unexpected business interruptions such as temporary shut downs (business interruption).

- Intended primarily for small and medium-sized business owners.

- Available and applicable to a variety of industries.

- The complexity of underwriting and claims, combined with the low volume and nature of transactions, pose obstacles to commercial insurance embracing InsurTech.

UCONN

Insurance data from originating insurer: XYZ Insurer

- XYZ Insurer's BOP historical loss experience from 2010 to 2020.

- More than 1.2 million data entries.
- Policy information:
  - Policy year, Earned exposure, Coverage limit
  - Exposure base: annual gross sales, annual payroll
  - Risk types: apartment, convenience store, fast food, motel, restaurant, retail, others
  - 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

UCONN.

InsurTech data from Carpe DATA / InsurTech innovations

CARPE provided us with real-time, dynamic data derived from emerging public data sources that shed light on numerous facets of a business: operations, products, services, physical plant, and many more.

{ CARPE DATA

- 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)

- Others

UCONN

- continued

Category	Explanation	Examples
Business Information	Basic information concerning the operation of businesses including address, year founded, operating hours, etc.	*coordinates, state, street_type, type_of_address, is_franchise, is_home_business, operating_year, language_spoken, opening hours*
Firmographics	Characteristics to segment prospect business including the type of business, the number of employees, the sales ranges, etc.	*business_size, credit_score, company_type, sales_range, revenue_range*
Classification	Categorization of a business	*category, segment, segment_details, NAICS code*
Risk Characteristics	Indicator features that identify current attributes relating to the potential risks of a business	*Alcohol, Chemical Application, Fireplace, Gasoline, Heavy Construction, Outdoor Heaters, Raw Seafood Served, Tobacco, Work at Heights*
Index	A suite of indexes on a $1 - 5$ scale targeting dimensions of risk that can be tuned by segment and location	*Customer Rating, Visibility, Reputation, Health & Sanitation, Maintenance & Condition.*
Proximity Score	Proximity scores identify risks associated with the surrounding businesses that may impose on policyholders in the immediate vicinity of that risk	*Negative Keywords, Combustibles, Entertainment, Traffic*
Territory Risk	Density scores of risks within a zip code area, engineered from collected information based on the location of a business	*TERRITORY.a1 to TERRITORY.m1 TERRITORY.a2 to TERRITORY.m2*
Text Data	(1) Webpage content crawled from a business' website; and (2) Customer reviews of a business from multiple data sources, including but not limited to the content of the review, star rating, number of likes, the date of posting, and the source of the reviews.	*Webpage content: content, title, url Customer reviews: content, likes, dislikes, stars, saves, language*

UCONN.

# Distribution of observed LOSS cost
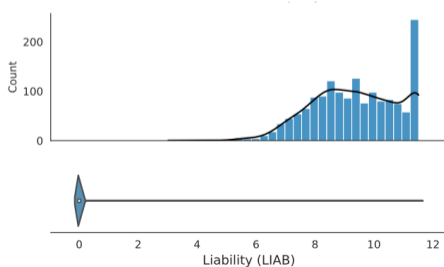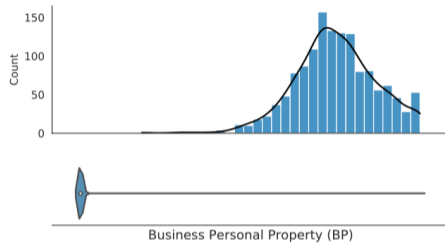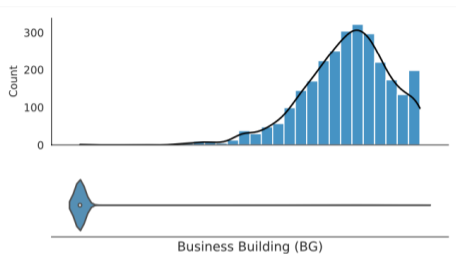
Risk type - BG coverage only
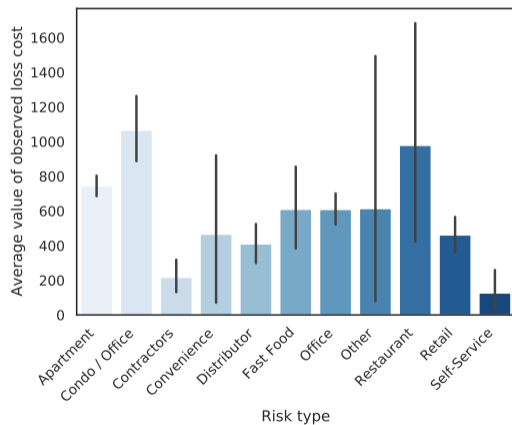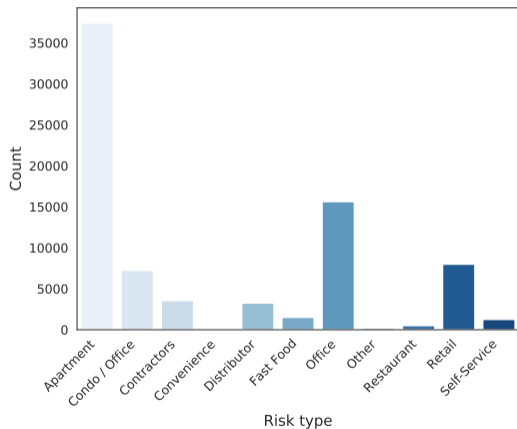
Model performance based on validation measures

| 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**

UCONN

Double lift charts

- Model lift refers to the ability to differentiate between low and high cost 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.

UCONN

Double lift charts: Building LGBM MAE Model
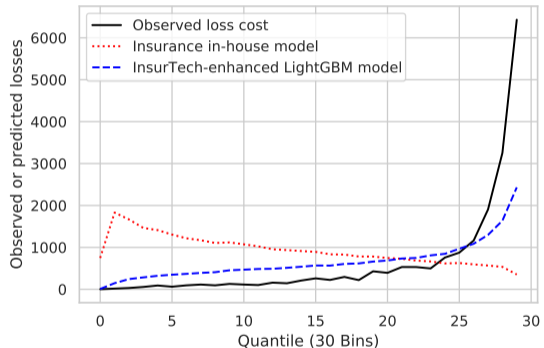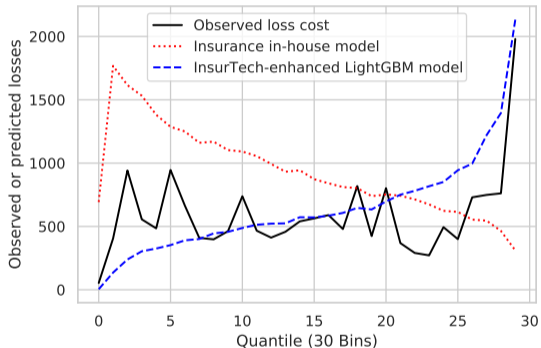
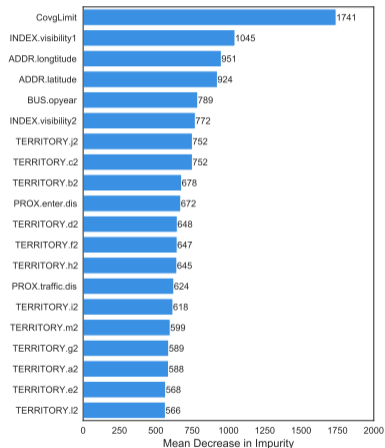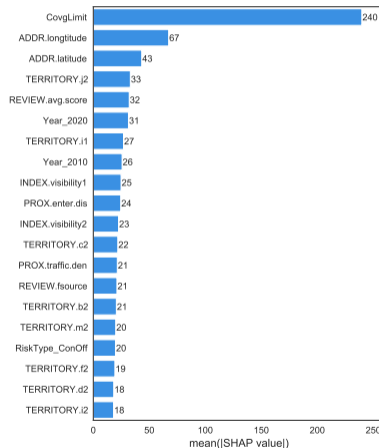BG Train   BG Test

Feature importance - Top 20 Features - Building


Mean Decrease in Impurity

SHAP values

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; and

(d) a business with no claim from the test dataset.

Illustrative cases   Influential features
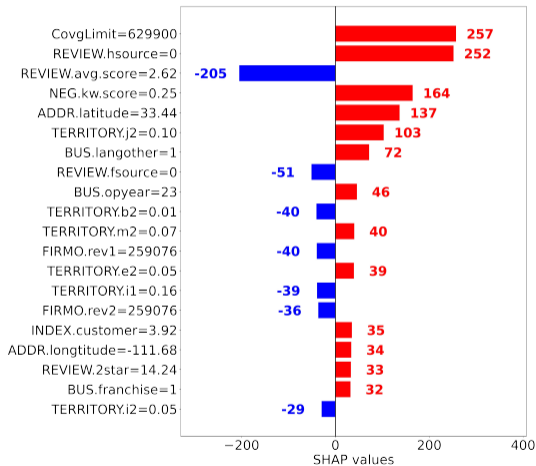
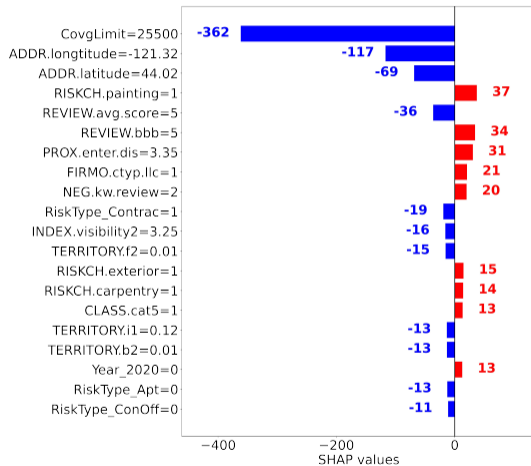# Top 20 influential features
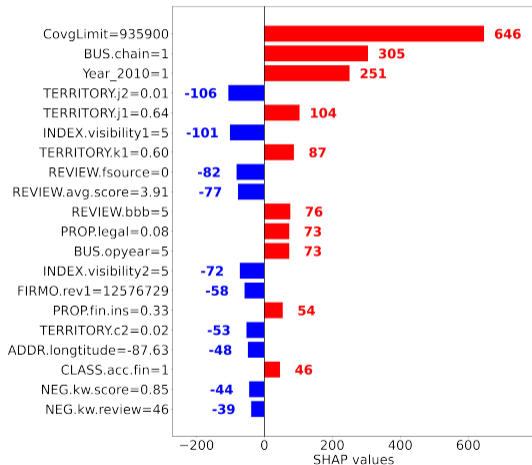
Business A

Business B

UCONN
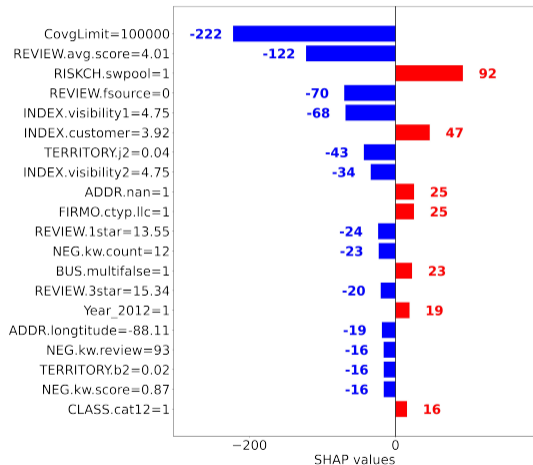
Top 20 influential features

Business C

Business D

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.

- Several ML algorithms were examined and we find that GBM with tree-based models as weak learners and using Bayesian optimization for hyperparameter tuning provides us a better benchmark model.

- This work is an example of the benefits that can be gained from a successful industry and university collaboration through the Illinois IRiskLab.

UCONN.