TECHNICAL DOCUMENT

PREDICTIVE ANALYTICS IN INSURANCE

Insurance is "a promise to provide compensation in the future if certain events take place during a specified time period" (source: www.casact.org). Unlike other products – whose cost is known before the product is sold – insurance is a different beast because insurance policies' price is unknown at the time of purchase. Hence, selling an insurance product carries a great financial risk.

At its simplest, a product price is defined as the sum of cost and profit. The primary aim, and biggest challenge, in the insurance sector is accurately estimating product cost. Over the years, insurers have developed a plethora of tools, methodologies, and mathematical models to calculate costs. The big data revolution – along with advances in data processing, predictive analytics, and artificial intelligence (AI) – has made this effort more achievable. Nevertheless, it's still difficult; in 2015, the U.K. motor insurance market made underwriting profit for the first time since 1994. This demonstrates how insurance remains a challenging business sector (source: abi.org.uk).

The key facts stated in the latest U.K. Insurance & Long-Term Savings annual report from the Association of British Insurers (ABI) confirms the importance of the insurance industry for the U.K.'s economic strength. The U.K. insurance industry is the largest in Europe and the fourth largest in the world with a total premium income of £300 billion in 2016. There are over 900 authorized general insurers in the U.K., which employ more than 300,000 people. The value of premiums written is constantly growing, with motor and contents insurance being the largest products. Over 75% of the U.K.'s households have had motor and/or contents insurance. Despite a total revenue in the tens of billions, fine margins and fraudulent claims totaling £800 million led to a £200 million underwriting loss in motor insurance.

The paramount objective in the insurance market is to set adequate, fair, and competitive premiums. With a customer-centric approach, an insurance pricing system should be easy to understand, provide stable rates over time, be responsive to economic drifts, and include loss control that provides affordable rates. These are very challenging and often conflicting requirements that place a large financial burden on insurers.

To provide premiums, insurers try to answer many unknowns throughout the customer journey (Figure 1). These include questions like:

- How risky is a customer?
- How do we retain more customers?
- How likely is it a customer will make a claim? Can we predict that claim amount?
- Can we identify fraudulent customers?
- How can we encourage customers to buy other products?

These are the main questions, but there are more as operations become more complex.

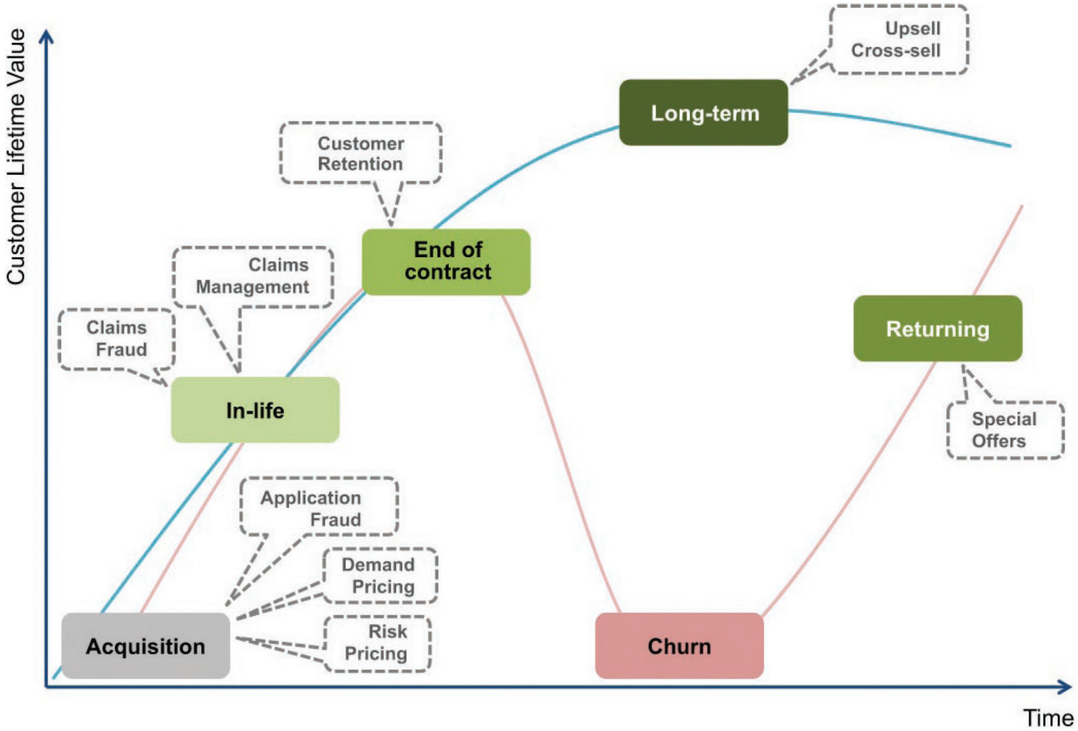

Figure 1: Customer Interactions Throughout the Customer Journey

Calculating adequate, fair, and competitive policies is the key to answering these questions and ensuring the long-term customer relationship. Hence, insurance pricing – often called ratemaking – is the key driver in the insurance industry and the art of data science within insurance. The two most important insurance concepts responsible for adequate, fair, and competitive insurance policies are pricing and claims. These concepts, supported by fraud detection, are the key analytics elements that create rapid insurance technology innovations.

Table 1 illustrates how data science can be utilized across these three insurance concepts and deal with various business challenges at different stages in the customer lifecycle.

Table 1: Leveraging Data Science for Insurtech

SEGMENT	CHALLENGES	ANALYTICS SOLUTION	TYPICAL MODELING APPROACH	BUSINESS BENEFITS
PRICING	The ultimate cost of an insurance policy is not known at the time of sale	Customer level Ratemaking (risk-based pricing)	Generalized linear models (e.g. the GENMOD procedure in the SAS programming language)	Adequate and fair pricing so the premium = loss + profit
	Understanding competitive market and its dynamics	Market-based pricing models including: conversion, demand and retention models	Propensity models	Expanding the customer base, competitive advantage
	How valuable are my customers?	Customer lifetime value	Survival analysis, segmentation, propensity models	Optimal marketing campaigns
	What is a customer pricing tolerance?	Price elasticity	Optimization	Maximizing profit
CLAIMS	Reduce high operational/ IT cost and maintain customer satisfaction	Claims management framework including first notification of loss	Holistic approach utilizing: propensity and regression models including bodily injury, claim cost and write-off	Real-time decision-making, monetization
FRAUD DETECTION	Application and claims fraud detection	Fraud detection framework	Holistic approach utilizing: propensity models, anomaly detection, fraud rules, black lists and link analysis	Minimizing loss

The successful development, implementation, and utilization of insurance predictive models relies on analytics platforms that must satisfy an extensive range of requirements including extract, transform, and load (ETL) capabilities; data manipulation, preparation, and visualization; model building and validation; model deployment; testing; production; and monitoring. Insurers often opt for a mixture of commercial and open-source tools to justify the implementation costs. However, careful consideration is necessary as this process can lead to suboptimal solutions since the integration process can be time- and resource-intensive.

Adequate, Fair, and Competitive Insurance Pricing

When looking for an attractive policy quotation, prospective customers usually ask, which insurance?In the very competitive U.K. insurance market, insurers need to develop a bespoke pricing methodology to ensure their policy premiums are adequate so they'd cover expected losses and incurred expenses are set up in a fair manner so the premiums are associated with expected losses and expenses, and competitive, so they attract new customers and retain existing ones.

Ratemaking or risk-based pricing is an essential step and the key element of insurance pricing. Actuarial pricing techniques and methodologies depend on insurance type, data availability, and extensive regulatory, marketing, and operational constraints. Furthermore, selected modeling techniques are constantly changing due to advances in technology and innovations in data science.

Techniques range from different statistical methods such as linear, additive and mixture models, to an extensive set of machine learning models such as random forest, gradient boosting, neural networks, or support vector machines. Statistical models typically work on a number of assumptions focusing on data fitting in the form of equations. In contrast, machine learning methods tend to make less assumptions about data and instead focus on learning through algorithm construction. Recognizing the most appropriate technique in the rich analytics landscape is often challenging and requires teams to consider multiple facets, including model accuracy, model stability over time, business constraints, required DevOps resources for model deployment and implementation, model response time, and so on.

Despite such model diversity, the Generalized Linear Model (GLM) remains the de facto standard in the insurance industry. Although machine learning methods can often achieve better prediction performance, GLM has gained popularity because it's easy to interpret and understand results, the model assumes a linear relationship between predictors and outcome, and it provides better control when selecting the rating factors. In addition, a GLM-based rating algorithm (Table 2) is easy to implement, fast to execute, and offers more flexibility to integrate actuarial experience and implement various constraints including regulatory, operational, and marketing constraints.

Table 2: GLM-Based rating relatives and rating algorithm (illustration)

Base Rate		£500	Rating Algorithm
Rating Factor	**Level**	**Relativity**	**Pure Premium = Base Rate**
Territory Zone	1	3.18	***Territory Zone**
	2	1.91	
	3	0.71	
	4	1.00	
Vehicle Age	1	2.38	***Vehicle Age**
	2	1.44	
	3	1.00	
Engine Power	1	0.37	***Engine Power**
	2	0.72	
	3	1.00	
	4	1.26	
	5	1.50	
	6	2.96	
	7	4.41	
Bonus	1	0.58	***Bonus**
	2	0.79	
	3	1.00	

The result of the ratemaking process is a predicted technical price (i.e., pure premium) that matches the likelihood of making a claim. If modeled accurately, it should provide the adequate price to cover expected losses. The technical price is then adjusted to include other underwriting expenses including acquisition cost, commissions, tax fees, and underwriting profit.

The traditional approach to insurance pricing has focused solely on the risk-based pricing, neglecting competitor prices. In the current climate, this approach is unsustainable, and many insurers are transitioning towards more sophisticated pricing methodologies. Altair encourages a holistic approach to insurance pricing illustrated in Figure 2. With a symbolic reference to ocean diving, the deeper you dive into a (data) ocean in a quest for discovering new (data) species, the more successfully you can seize new, valuable (business) gems.

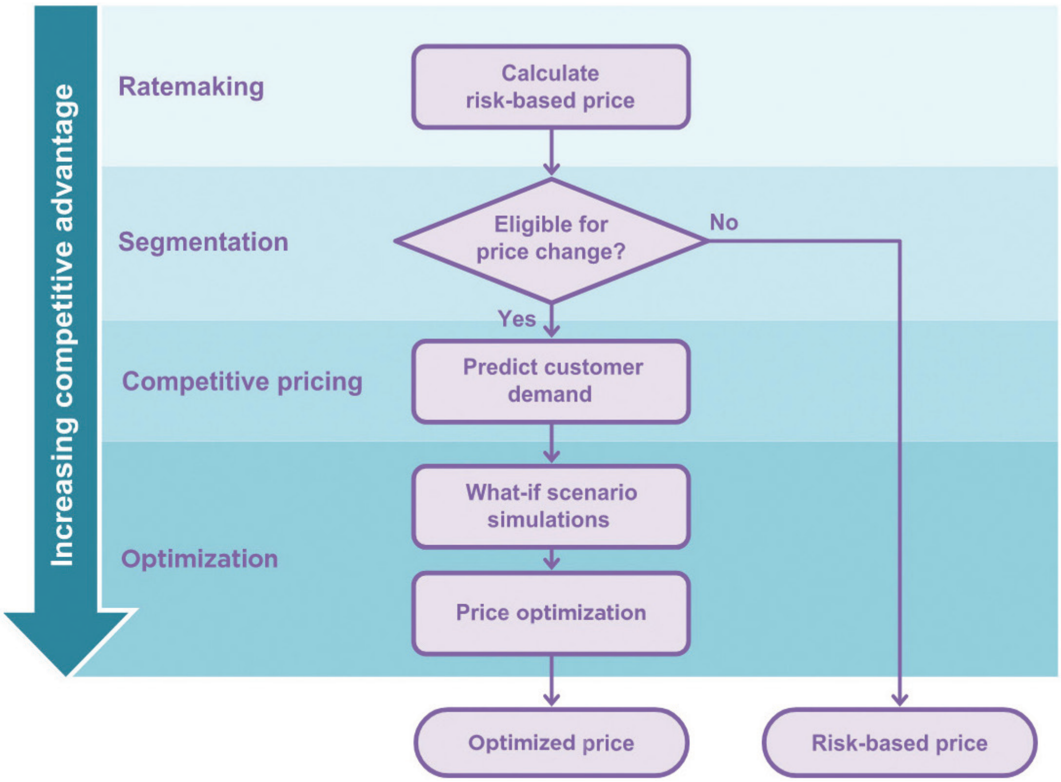

Figure 2: Insurance Pricing Process

In this analogy, premiums based solely on risk-based pricing would be like searching in a shallow water, which wouldn't let teams maximize the data's full potential. Ratemaking pricing ensures an adequate price, but not necessarily a competitive one. Hence, the pricing process should comprise the additional elements, including customer segmentation, competitors' prices, and price optimization.

Customer segmentation is an important step to identify if a customer should be targeted for price increase or decrease. Additionally, it serves as a reassurance that premium discounts are only being offered to the "safest" customers. Segmentation can be very simple based on a few business rules, or more sophisticated based on a propensity model, such as the probability of making a claim or a clustering model that creates segments by eligibility levels.

Competitive pricing is about adjusting the risk-based premiums to competitors' rates. Depending on marketing initiatives, demand models may be required to capture the competitive market – a conversion model for customer acquisition, or a churn model for customer retention. The decision as to which binary classifier to utilize for building these propensity models will depend on many factors, including the required model performance and ease of model deployment and implementation. In addition, conversion models usually run in real-time – which is why model scoring speed is an important factor to consider when deciding on modeling technique.

The final step of the pricing process is price optimization, usually referred to as price elasticity and is about assessing price tolerance at an individual level. This step provides extra benefit ensuring the relevant competitive model (the yellow line in Figure 3) sets up the price, which maximizes the profit (the blue line in Figure 3). An optimization model usually requires price simulations for different what-if scenarios so the maximum benefit can be extracted given the optimization constraints.

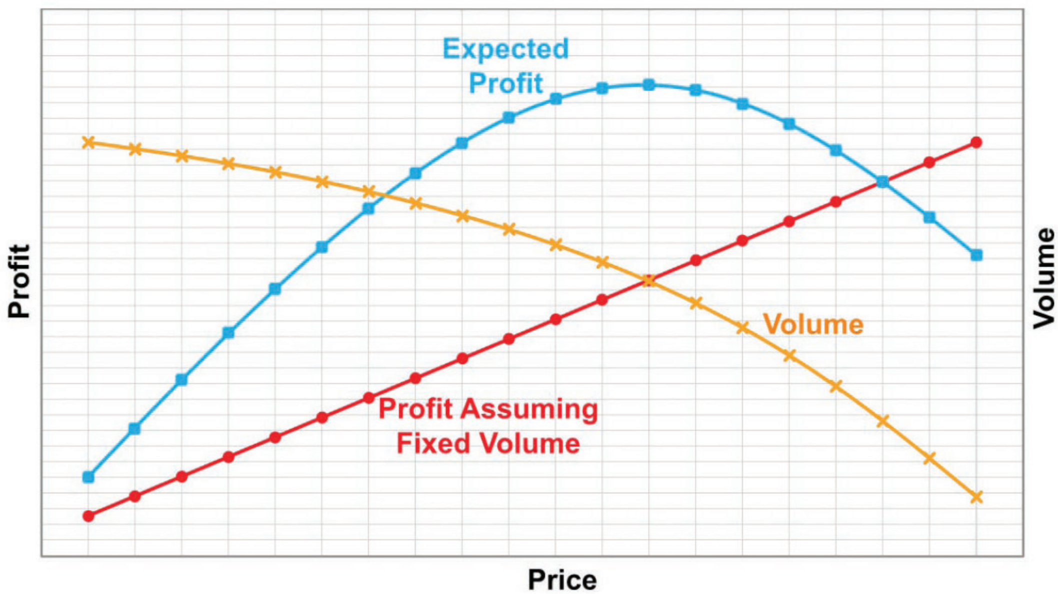

Figure 3: Price Optimization Problem

Figure 4 is an illustration of an optimization model that offers premium discounts of up to 20% on the risk-based price. Depending on the underlying demand model, the premium discount distribution ranges from no-discount to a maximum 20% discount.

Once the optimal scenario is selected, the complete pricing solution is ready for deployment, implementation, and testing. This includes the risk-based model, the demand model, and the optimization model. The complete model suite can be implemented on a single rating engine or hosted across multiple engines. After a rigorous testing process, the pricing solution is ready for a racing competition.

Claims Management

An insurance claim is "a formal request to an insurance company for coverage or compensation for a covered loss or policy event" (source: www.investopedia.com). Once initiated, the claim often goes through a complex process with one of two possible outcomes – the claim is either accepted, leading to a settlement, or it's rejected. The claims process would typically be contact the insurance company, start the claimant investigation, check the policy coverage, evaluate the damage, and arrange compensation.

U.K. insurance industry figures are staggering. On average, insurers dish out £33 million per day in motor claims, £13 million in property claims, £12.5 million for policy protections, and £1 million for travel claims; the average bodily injury claim is close to £10,000; more than 98% of motor claims have been accepted; and the yearly cost of fraudulent claims is £1.3 billion. Such massive claim expenses can lead to an underwriting loss, this is especially evident in motor insurance where an underwriting profit has only been made once in the last 24 years. (Source: www.abi.org.uk, 2017)

Clearly, insurers are faced with a number of challenges including high operational cost, constantly increasing customer demand, increased fraudulent claims and a lengthy process, hence customer dissatisfaction. Additionally, high IT cost, lag in change request, and poor IT/third-party integration increases operational costs, which leads to underwriting losses.

Despite the ongoing efforts of improving claims processes and fraud prevention, there's room for significant improvement that focuses on better customer service and customer experience, improving operations, and managing the claims more effectively in terms of both time and resources.

To achieve these improvements, we must integrate systems as best we can, and continuously incorporate advances in predictive analytics and ambient computing such as GPS car tracking, telematics devices, body activity tracking, and image recognition. Machine learning and AI can deepen retrospective analysis and ensure decisions are informed by data, not subjectivity.

One of the important areas of InsurTech innovations is support at first notification of loss (FNOL). FNOL, as the very first step in the insurance claims process, is often deemed a bottleneck of the process where a claims adjuster is faced with a number of challenging and sensitive issues, including the possibility a fraudulent claim, policy coverage, and loss assessment. Failure to deal with these issues in a fair, effective way can damage customer relationships.

An FNOL decision support system can assist in this complex, time-sensitive decision-making process using optimal resources. The FNOL ecosystem is a real-time solution that utilizes a range of predictive models, AI software, social and third-party networks, and ambient intelligence.

Since predictive models are typically seen in motor insurance, models are deployed on a scoring engine as RESTful web services and connected to a front end. They are simultaneously scored in real-time during the first notification of loss, and model scores are visualized on the dashboard.

The choice of predictive models depends on policy type and insurers' preferences. Motor insurance models typically include predictions of a vehicle being written-off or recovered; a bodily injury severity score would assist in more accurate predictions of the estimated claims reserve amount; total settlement cost would give total estimated claim cost; and a fraud model would flag potentially fraudulent claims. Additionally, the fraud model could be designed using an existing company's fraud methodology and implemented as a composite solution consisting of a series of models and fraud indicators.

Represented on an intuitive dashboard, the FNOL solution acts as a handrail that helps claims adjusters better allocate resources and prioritize tasks. Decisions based on the dashboard indicators, such as an early settlement cost offer, can minimize unnecessary administrative costs that teams would've otherwise incurred.

Teams can also implement dynamic questionnaires to direct the claims process by creating a bespoke strategy based on the dashboard scoreboard (model outputs). For example, if the fraud gauge is “flashing red,” then a different set of questions can be pulled out from the database to probe the claimant. On the other side, a “steady green” light would skip additional questions, making the claims process faster – which increases customer satisfaction.

With the overall tendency towards more customer-centric services, the cutting-edge InsurTech innovations are focused to self-service solutions available anywhere, anytime. The latest FNOL tools utilize automated insurance agents (known as claimbots) that make conversation, exchange information, and make assessments and recommendations faster than humans.

Even though it’s been around for several years, the potential of the FNOL decision support tool hasn’t been fully exploited. A fraction of insurers utilize the tool in its simplest form. Market research shows that insurers are willing to adopt AI technologies; however, implementation and integration costs, data security considerations, and imposed regulatory rules have been obstacles in implementing these technologies.

Given how urgent it is to make the claims management process more effective, the demand for such support tools is evident. The promising news is that teams and organizations can take baby steps that lead to a better, more efficient claims management decision support ecosystem.

Request a Demo:
altair.com/contact-us