

# Predictive Analytics in Insurance: A Thought Paper

N. V. Subramanyan

## Contents

*Introduction*

*Application areas for PA and ACRM*

*Premium Fixing*

*New Business Underwriting*

*Non-Life Insurance*

*Life Insurance*

*Lapsation*

*Sales Force Relationship Management*

*Claims and Fraud Prevention*

*Conclusion*

*About the author*

## Introduction

It is questionable if Ralph Waldo Emerson's often quoted remark, 'Build a better mousetrap and the world will beat a path to your door', is totally true anymore. While that statement is in favor of innovation and could be generally true, the insurance market is now so competitive that mere innovation may not be sufficient anymore to maintain profitability and remain competitive. What with the intense competition insurers face in retaining their customer base, complacency and sticking to the old order of working is fraught with danger today. Rather, a radical rethink is the need of the hour.

Indeed, Predictive Analytics (PA) and Analytical Customer Relationship Management (ACRM), which is a recent commercial application of Predictive Analytics, are ways by which insurers can try to be ahead of the innovation curve. PA encompasses a variety of techniques from statistics, data mining and game theory to analyze current and historical facts to make predictions about the future. It exploits patterns found in historical and transactional data to identify risks and opportunities allowing models to capture relationships among many factors. This enables the assessment of risk or potential linked with particular sets of conditions, thus guiding decision making. PA can be used in actuarial science, financial services, insurance, telecommunications, retail, travel, healthcare, pharmaceuticals and other varied fields. Perhaps the most well-known application of PA is in credit scoring, used in financial services, which amalgamates customers' credit history, loan applications, customer data, etc., to rank the likelihood of future credit default.

Owing to the nature of the business, customer retention is very important for insurers, more so during and after a recession. During recessions, buyer confidence is low, which means keeping the customers you have is even more important than attracting new customers. Further, with consumers being more tightfisted, existing customers are much more valuable than ever. In addition, the cost of retaining existing customers is much lower than that of attracting new customers. Insurers could use ACRM as a business strategy designed to optimize profitability, revenue and customer satisfaction, by developing the entire enterprise around customer segments, enabling and encouraging customer-centric behavior and processes. ACRM technologies support this strategy by providing greater insight into customers and their behavior through improved management of customer information and more effective customer interactions. PA is another area that insurers should consider investing in to be able to increase their retention levels.

Though insurance depends on the law of large numbers for effective risk dispersal and to be successful, it is strange that throughout history this industry has globally been quite product-focused rather than relationship-focused. Perhaps, the thinking so far has usually been that there is no need to have regular contact, since for most buyers, once the premium is paid, they do not really connect again with insurers until it is time for renewal, or if they need to notify them about claims.

However, PA and ACRM are potent tools for insurers to address the needs of their sales force and end customers by targeting them on the basis of their potential, and making their marketing programs more focused and productive. For instance, insurers can sell better products to their customers when they have better information on their requirements, via new offerings, additional riders or through identifying cross-sell and up-sell opportunities. Further, insurers can begin to look more like diversified service providers, and the sales force might be able to market other products on offer. This means that the gulf between insurers and customers will diminish considerably. Optimizing product offerings is the goal of every insurer, and it can be realized only through better understanding of the customers' needs. This can only be achieved when insurers have a closer view of the market that CRM and PA enable. Additionally, since PA and ACRM can help to identify potential lapses, frauds, etc., insurers can take corrective steps proactively.

Thus, the chief advantages that PA and ACRM offer would be:

- ? Increased customer penetration by getting more out of established customers
- ? More effective use of resources through identifying and concentrating on the best customers, targeted customer development efforts and keeping the existing customers loyal
- ? Achieving greater efficiencies by allowing users to work smarter and not just harder

## Application areas for PA and ACRM

While there are numerous areas where PA and ACRM can be used by insurers, the following areas stand out for immediate address owing to the impact they could provide:

### Premium Fixing

PA and CRM can help in appropriately pricing the insurance products and specific risks so as to avoid pricing deficiencies, like underwriting leakage due to charging lower premium for a higher level of risk, or loss of business due to charging higher premium for low risks. PA and CRM can help insurers to accurately assess risk through better prediction of claims incidence and severity, thus improving pricing efficiency. Accurate predictions enable optimal reserving and profitability.

### New Business Underwriting

The primary insurance application for PA and ACRM is in

underwriting, which essentially deals with decision making on risk acceptance. The major problems insurers face today are regarding customer behavior and the changing priorities of customers. Such problems include:

- **Increased policy shopping:** With the availability of price comparison sites or insurance aggregators, and the trying economic conditions that prevail, today's consumers shop for discounts in insurance. Hence, getting and retaining customers is very challenging today, though it is also more crucial than ever.
- **Fewer assets sold or insured:** Decreased economic activity has led to fewer new policies being sold, which is challenging bottom lines.
- **Increased coverage levels:** Customers may feel the need to increase coverage as a hedge against disasters during times of reduced wages or unemployment. Paying a slightly larger amount is less of a hardship than having to pay an exorbitant amount suddenly due to inadequate coverage.
- **Need to optimize resource utilization:** It makes sense for insurers to utilize their resources in the most optimum manner, and therefore, to focus them on the most profitable areas.

Due to the above factors, it has become crucial for insurers to select and price their risk appropriately so as to avoid anti-selection and ensure that their portfolio remains healthy. The most important questions that insurance underwriters face are as under:

## Non-Life Insurance

- Which risk has the least probability of resulting in a claim?
- What additional risk mitigation strategies will prove helpful for this particular cover?
- Does the geographical area of this risk have an unduly large bearing on the outcome of the event?
- Does this class of risk deserve preferred category status?
- Is there any under-insurance suspected or observed?
- On which risks should the limited underwriting time be invested?

## Life Insurance

- Which applicants have the best chance of remaining healthy during the policy term?
- Which additional tests/data will prove unnecessary for a particular case?
- Which lives truly deserve preferred category status?
- On which of the applicants should the scarce underwriting time be invested?
- Designing a battery of new tests based on the profile eg. probably ELISA for policies originating from a location with high incidence of AIDS

Present underwriting methodologies, though useful in their own way, could be inadequate to deal with changing customer risk

profiles and in meeting new challenges—including the regulatory changes—in the market. PA and ACRM provide another way to address the situation without the need to invest the same amount of time and money. However, it would be essential to expand the available data elements from both internal and external sources without further inconveniencing the customers. Such elements could be as below:

Life Insurance	Non-Life Insurance
<ul style="list-style-type: none"> <li>• Savings background</li> <li>• Place and length of residence</li> <li>• Previous insurance purchase trend</li> <li>• Daily commuting distance</li> <li>• Designation at work</li> <li>• Foreclosure/bankruptcy indicators</li> <li>• Marital status</li> <li>• Reading habits</li> <li>• Growth at work</li> <li>• Fast food consumption habits</li> <li>• Availability of health facilities, etc.</li> </ul>	<ul style="list-style-type: none"> <li>• Previous claim experience</li> <li>• Place of business and disaster exposure</li> <li>• P r o p e n s i t y f o r u n d e r i n s u r a n c e / o v e r - i n s u r a n c e</li> <li>• Employee unrest, etc.</li> <li>• Environmental factors</li> <li>• Relationship with regulators, etc.</li> <li>• CSR activities, etc.</li> </ul>

Algorithmic solutions built from the above and many other such data elements can help to quantify the acceptability of individual applicants in life insurance and the nature of risk undertaken in non-life insurance in a better manner than traditional underwriting requirements. And based upon these, a scorecard can easily be built for these cases, which would greatly ease decision-making. A sample scorecard can be as under:

Data	Traditional U/w Data	U/w Model	Purchase Likelihood Model
	<i>Application</i>		
Demographics	X	x	
Medical History	X	x	X
Education	X	x	X
Family History	X	x	X
	<i>Medical Exams</i>		
Height/Weight	X	x	X
EKG etc.	X	x	X
	<i>Financial Aspects</i>		
Loans etc.	X	x	X
3 <sup>rd</sup> party data	X	x	x
Decision	Good risk	Possibility of lapse	Possibility of early claim

Source: Fictitious data, for illustration purposes only

Life Insurance PA and ACRM Model

Sources of such external data could be:

- Survey or medico-actuarial models
- Past experience of industry
- Observed lifestyle data

- Banking industry for income level, net worth, savings, loans and investments
- National census, etc.

**Lapsation**

For life insurers, policy lapsation has perennially been the biggest challenge, and it has been a growing concern, especially in emerging markets, due to the pace of new product development and competition for policyholders' investment money. Early-stage policy lapsation, perhaps prompted by mis-selling and other unhealthy practices adopted by the sales force, and also the intrinsic inertia on the part of customers towards buying or continuing insurance, severely affects retention, and thereby, the bottom lines.

PA and CRM can generate predictive lapse risk scores for policies or groups of customers to highlight their likelihood to lapse, by using data from different dimensions, such as policy line, customer profile, sales personnel interaction and other transactions. Further, VAR techniques would be useful in analyzing policies and focusing collection efforts on high value or high risk policies.

**Sales Force Relationship Management**

Like in any other business, the insurers' relationship with the agents and other sales force is crucial. PA and ACRM solutions can be used at different stages in the distribution lifecycle to improve productivity, monitor business quality and manage sales force development training needs. The various stages in distribution where this can be used are:

- **Sales force prospecting:** Insurers could focus attention on targeting the right agents for appointment by indentifying their potential with appropriate metrics analytics.
- **Mentoring and monitoring:** PA and ACRM help to estimate average agent lifetime value and use it to identify high-potential agents early on in their tenure, so that they can be properly trained and empowered.
- **Sales force retention:** With the help of PA and ACRM, insurers can identify potential agents with risk of attrition, so that they can target retention efforts to prevent this.
- **Risk management:** Insurers can predict an agent's tendency to mis-sell, and thereby take steps to mitigate this risk and ensure quality business.

**Claims and Fraud Prevention**

Traditionally, claims have been an area of concern for insurers, with fraudulent claims being observed from the early history of the industry. And with the advent of technology and regulatory protection, such instances of fraud have also increased manifold. Insurance businesses have also traditionally been susceptible to internal fraud. Internal fraud occurs when an employee exploits the vulnerabilities of the system to route benefits to either his or her self or an accomplice. Historical trends show an increase in

internal fraud during economic crises. Unfortunately, unlike external fraud, internal fraud is very tough to detect or prevent at early stages.

Typically, some lines of insurance such as medical insurance or auto-insurance could be more vulnerable to fraud. In life insurance, fraud is more by way of nondisclosure and obtaining cover when an adverse health condition is already diagnosed. A rule of thumb is that approximately 90% of insurance fraud is the result of claims padding by adding damage, injuries and fictitious persons to claims. Another 10% could be by organized event staging, fabrication of documents, etc.

Frauds cost the insurance industry huge amounts of losses annually, and with claims investigation being a laborious process and prone to criticism from customers, regulators and the press, very few claims are actually verified by internal or external investigators. This is also mainly due to the operational burden involved in the investigation of claims. With the regulatory focus being on insurers for speedy settlement, a very small percentage of processed claims are investigated annually, and a smaller fraction of these are proven as fraudulent.

However, the silver lining is that PA and ACRM, with the help of data-mining techniques, can leverage these historical instances of fraudulent claims and funnel out those with higher chances of being fraudulent, thus prioritizing claims for detailed investigation. Modeling for external fraud can be a pretty straightforward exercise, but the modeling of internal fraud can be much more complicated—especially due to the rarity of its getting caught. While external frauds may be more frequent, they may not leave enough evidence to be modeled, while internal frauds provide enough evidence but may not be too many in number. Thus, a balance may need to be struck. Over-sampling, weighting or other suitable statistical techniques may be chosen to address these problems appropriately.

With suitable data elements being available, it would be possible to help insurers with the identification of suspected fraudulent claims. These data elements may be considered together with other suitable analytical approaches to monitor these unusual patterns of claims.

Of late, the global insurance industry has evolved towards better recording and organization of data pertaining to customers, employees and agents. Transactions related to shopping, policy endorsements, claims and retention, including service level touch-points, are now generally part of the insurers' database. In particular, detailed claims data is captured right from the very first notice of event to settlement. Some such data elements available with insurers are:

- Demographic details, such as date of birth, address, occupation, etc.

- Claim details, such as time of event, type of loss, quantum of loss, amount claimed and settlement amount, total number of such claims made and settled, etc.

Data preparation to model fraud is the first step towards translating an instance of fraud into its component data elements. The above data elements could be further transformed into:

- Geographic distance between claimant, hospital and insurer
- The total number of supplements for a claim
- Difference between the allowable payment amount as against actual amount paid
- Number of claims concentrated in a single location or region. For instance, is there an unusually high number of auto claims being made from the same locality, with similar reasons being given?
- Too many fire claims coming from a region where there is no corresponding record with the fire authorities
- Multiple claims to the same or similar sounding parties in close proximity or during the same period of time
- Several health claims originating from the same hospital, or many auto repair bills coming from the same mechanic
- Too many death claims with similar reasons, arising frequently from the same locality, etc.

Other similar attributes can be suitably captured to define different instances of fraud for modeling. Understandably, classifying the target attribute is the key to effective modeling. A popular data-mining technique to predict unexpected or unusual outcomes is Binary Response Modeling or Logistic Regression, whose goal is to forecast the occurrence of fraud at each individual instance, based on the data elements mentioned. Generalized linear models may also be applied to forecast fraud.

It must, however, be noted that in any approach, only claims investigated and proven as fraud may be used to identify key characteristics and predict future frauds, unless the modeler is able to imagine and identify other potential areas of fraud. So, validation is important, and appropriate re-sampling techniques for model validation, like bootstrapping and jackknifing, may be used to improve model accuracy.

However, the above approaches may be incomplete, as they may not consider uninvestigated cases as clean, while there could be fraudulent claims that could have been identified if investigated. Inputs from domain experts to build unsupervised learning techniques into the algorithms, along with PA and ACRM models, and social networking or visual link analysis would be helpful. The performance of the model needs to be monitored and revised continuously, and efforts to investigate random cases should continue to combat new fraud patterns.

The analytical approaches discussed aid in the detection of internal fraud. But the key to the success of any model lies in understanding the business challenges and adopting the right analytical

approach, coupled with domain expertise. Based upon these, a scorecard can easily be built for these cases, which would greatly ease decision-making. A sample of this can be as under:

Data	Traditional Claim Data	Claim Model	Fraud Likelihood Model
	<i>Personal</i>		
Demographics	X	X	X
Medical History	X	X	X
Education	X	X	X
Family History	X	X	X
Marital History	X	X	X
Previous Claims	X	X	X
	<i>Occupation</i>		
Employer	X	X	X
Salary drawn	X	X	X
	<i>Financial Aspects</i>		
Loans etc.	X	X	X
3 <sup>rd</sup> party data	X	X	X
Decision	Clean claim	Possibility of enhanced claim	Possible fraud

Source: Fictitious data, for illustration purposes only

Life Insurance PA and ACRM Claims Fraud Identification Model

## Conclusion

It is to be noted that insurance companies have been slow to accept and use PA and CRM techniques so far. However, considering that they are now operating under tight budgets, and soliciting new customers and retaining existing customers is extremely challenging, PA could hold promise in their area. While some difficulty in budgeting may be seen, it is expected that timely investments in this area would be helpful in the long run.

## About the author

N. V. Subramanyan heads the Insurance Practice at NIIT Technologies. He has over 22 years of experience, with over 16 years in the Insurance Domain and about five years in Management Consultancy. He is an ACII from the Chartered Insurance Institute, London.

For more information, visit at [www.niit-tech.com](http://www.niit-tech.com) or contact [it.solutions@niit-tech.com](mailto:it.solutions@niit-tech.com)