



Predictive Analytics in Insurance Services



Ashok Goudar

Associate Vice President & Senior Architect, Europe
Mphasis

Contents

1. Introduction	4
2. Large data challenges in data mining	5
3. Perception of complexity	6
4. How to simplify the perception of complexity	6
5. Competitive advantages	7
6. Value to customers	7
7. The role of IT strategy	8
8. Conclusions	8



1. Introduction

Predictive analytical techniques such as risk, affinity and customer behavioral score cards, are widely being used in finance, payments and retail industries today. So far, adoption in the insurance industry is at its infancy stage, most often confined to large insurers. Predictive analytics are mostly used as an extension to existing MIS/BI analytical reporting solutions but now predictive analytics is gradually finding application in multiple areas such as policy risk scoring, claim fraud detection, subrogation, referral scoring, and score based customer segmentation for marketing and service purposes.

Another reason for increased adoption of predictive analytics is because of the increased level of adoption of data warehouse and BI/MIS reporting solutions in many insurance organizations, where such existing DW/BI capabilities can easily be extended to adopt more effective predictive analytical solutions. The recent increase in availability of large data (Big Data) from a variety of data sources has opened a new dimension in predictive analytics space, enabling the insurers to take advantage of such large source of data, along with its existing traditional data, to gain useful insights that can reap commercial, business and IT benefits to the insurers and to their customers.

2. Large data challenges in data mining

In insurance organizations, historical data pertaining to policies, claims, customers, partners and market data plays a significant role in building highly effective predictive analytical models such as policy risk scoring models, claim fraud risk scoring models, subrogation models and customer churn models. As the accuracy of the key coefficients and factor variables depend upon the underlying data (to train the model), higher the volume of the relevant data, the more accurate will be the derived predictive models (through model training). Also, the precision of the data mining outcome increases with the growth in size of the right target population.

In addition to structured data, a large quantity of unstructured data (Big Data) is originating by a variety of relevant external sources such as twitter feeds, web/internet blogs, and machine generated data. Such data of interest, which is usually unstructured, can be used in conjunction with structured data for data mining and predictive modeling to derive useful and highly critical insight about various aspects of the business and customers. Organizations often maintain historical transaction data (claim, policy, customer data) at least up to 10 years and such structured data, along with other unstructured content, can be of enormously large size, in the order of multiple terabytes or even petabytes. Processing such large quantities of data during data mining, can pose some of the following challenges:

Enormous data preparation effort – To apply correct statistical data mining methods such as linear, nonlinear, log-linear, classification/clustering (k-factor), regression, neural networking including decision trees etc, the data needs to meet structure as required by the chosen model. Data preparation effort would usually include the following, to make the target data to be fit to model with the chosen model, to improve the accuracy of the mining:

- Data type conversions and transformations – Multiple data mining models require dependent and independent variables to be numeric or categorical (in some cases). Such data type conversions can take enormous effort with large data sets. Converting categorical (string) data variables to binary (0 or 1) data type can take enormous data preparation effort.
- Missing value analysis – Missing or incorrect values in the data can directly impact the data mining results. Before applying the data mining models, it is necessary to perform missing value analysis using techniques such as missing data pattern identification, estimate missing value statistics such as means, deviations, correlations (pair wise variables) and fill the missing values with estimated values using regression, estimation and multiple imputation methods. Large data sets can take enormous time and effort to perform missing value analysis.
- Higher levels of cleansing and noise reduction effort– Higher the volume of the data, larger will be volume of the incorrect or missing data (noise), requiring expensive cleansing and noise rectification processes.
- Target data set (Universe Population) preparation – Large volumes of data (Big Data sets) can impose new constraints when defining training (development) data sets and testing data sets, that can truly represent the underlying population universe. Also to improve the accuracy of the prediction, it is important to run the predictive algorithms against the complete universe of the target data, in which case, the extremely large data sets can induce performance issues.
- Performance of the mining algorithms – Using traditional data mining tools and platforms, large volumes of data can impact the performance of mining processes, taking considerably higher time for the execution of mining algorithms, almost making it impossible to process against large production data sets. The performance can become even more important with non-linear and mixed models, which use multiple iterations to improve the statistical estimates in the models.

3. Perception of complexity

Real-world insurance systems are complex and continuously adopting systems, constituents of which have multiple interactions and connections with various agents and subsystems, in which relationship between the predicted and factored variables are often non-linear and can not be accurately predicted by applying simpler linear models. By (over) simplifying the model to fit linear statistical models, such as linear regression analysis, often important and useful information in the system can be lost. Such complexity in the systems, which often exhibit self-learning and constantly adaptive features, can be best explained with non-linear statistical models, such as neural networks, clustering, decision trees methods and machine learning procedures.

The complexity in the system is handled by training the model with historic data, with which the model improves its statistical factor estimates (coefficients of independent variables), and is subjected complexity simulator models, using which the model predicts the future outcomes.

Predictive analytical models are considered to be complex to model, implement and analyze, as they require specialized statistical and mathematical skills to develop the models and analyze the outcomes.

The traditional predictive analytical tools and software packages would still require a considerable degree of statistical and mathematical knowledge, to practically use such tools. This makes it difficult for the insurance business users, marketers and even to actuaries, to use predictive analytics quickly into their business portfolios.

Existing predictive analytical tools (which are mainly statistical tools) require steep learning curve, to effectively use such tools to solve business problems.

Real time predictive analytical solutions dealing with large quantities (Big Data) (terabytes, petabytes) of structured and unstructured data are, by and large, considered being very complex to model due to the problems associated with data formats as text, video, image etc., to be used with structured data present in rows and columns and also due to security and data privacy issues associated with “Big Data”.

4. How to simplify the perception of complexity

The complexity in the model increases with the number of the independent variables (factored variables/agents) and their non-linear relationships, with predicted variables (claim risk score, policy score etc.) and co variables. Usually, higher the number of variables, the larger will be the complexity of processing the predictions. For instance, propensity for claim fraud can depend up multiple factors such as age, gender, location, country, insurance product, past history etc.

The complexity in predictive analytical models can be simplified by carefully analyzing the models and applying various dimension reduction techniques, to simplify the models, before taking them forward for further analysis. Some of the following types of dimension reduction techniques are helpful to simplify the models:

- Factor analysis – Techniques can be used to analyze and screen the variables, and choose the right set of variables that can explain higher degree of relationship with the predicted variables and thus simplifying the complexity in the model.
- Correspondence analysis – To analyze the relationships between the variables to select the right set of variables for further analysis.
- Scaling techniques / multi dimensional visualizations - are used to determine (visualize) the perceptual relationships between the objects and thus to select useful variables, that can be used further in a predictive analytical model.

Complexities associated with large data in training, validation data sets, can be addressed by using structured data sets (samples) that are representative of complex design requirements of the underlying model, so that the result outcomes are valid. Various complex data sampling techniques such as clustering, multi staging sampling, stratified sampling, non random sampling, and unrestricted sampling etc. can be used, to carefully choose the training, validation data sets, to improve the validity of the results.

The perceived level of complexity of predictive analytics can be further simplified by leveraging to latest BI and predictive analytical tools and software packages, by cross training the statisticians and mathematicians (actuaries) into business and marketing functions and vice versa, and by systematically institutionalizing the predictive analytics, across all levels of users in the enterprise.

In some cases, it might become impossible to statically establish meaningful relationship between the chosen factor variables and the predicted variables, in such cases, the existing insurance knowledge among the users (underwriters, claim handlers) can play a big role in modeling the right perceived association (based on experience) between the variables, thus reducing the complexity in the model.

5. Competitive advantages

Adoption of predictive analytics can help organizations to have competitive advantages in terms of cost, customer relationships (customer satisfaction), and market leadership, on account of the following:

- Predictive analytics will empower the insurers to understand their customer profiles accurately, which can be used to provide highly tailored and personalized services across sales, service and marketing channels.
- The claim handling processes are much simplified and insurers can provide fast track claim settlements for low risk claims, directly resulting in increased customer satisfaction and significant savings in claim handling costs.
- The policy risk scoring will enable insurers (Actuaries) to accurately determine the premium rates, which can result in improved overall profitability of the business.
- A well modeled predictive analytics can also help the insurers to accurately determine the claim subrogation success model and hence reducing the overall liability of the claim (by increased subrogation recovery), which can result in increased profitability. Such models can also help to improve accuracy of claim reserves.
- Most importantly, the predictive analytics can reduce the operational costs, in policy cycle management and in claim handling operations (loss to expense ratio), and such savings can be passed on to the customers, in terms of lower premium rates and faster claim settlements.

6. Value to customers

Predictive analytical solutions bring value to both the customers and also their insurers. From the customer point of view, predictive analytics can offer following direct benefits:

- Simplified claim handling processes – through reduced inspections, queries, referrals etc.
- Faster and automated, claim settlements for low risk claims – drastically reducing the settlement times.
- Reduced policy premiums for low risk customers.
- Improved customer services, during policy administration and claim handling cycles.
- Faster rehabilitation and associated claim services.
- Reduced policy premiums due to reduced operational costs at the insurer side.
- Highly personalized and preferential services to the customers, during the entire life cycle of the policy and during claim settlement services.

7. The role of IT strategy

IT plays a critical role in adopting practically feasible predictive analysis solutions in an insurance organization. The role of IT is evident in all stages of predictive analytical processes. The following are the key areas, where role of IT is highly crucial in adopting effective predictive analytical practices:

- IT plays a critical role to facilitate tools, software package and solutions for data mining and predictive analytics that can be easily be leveraged by business users.
- IT can play a big role is providing solutions and infrastructure, which can process very large quantities of unstructured data, emerging from new sources such as web click streams, twitter, and internet blogs, that can provide very useful information relating to various aspects of the business.
- Data quality plays a critical role in determining the accuracy and effectiveness of the predictive analytical models. IT must a play a strategic role in maintaining the data quality in down stream and upstream systems.
- Enterprise level predictive analytical processes are highly dependent upon the detailed and summarized high quality data held in purpose-built data warehouses. IT can play a big role in deploying high performance, scalable data warehouse/BI and next generation predictive analytical platforms, built on MPP, embedded analytics, distributed data warehousing and hardware based ETL solutions, that can support real time or near real time predictive analytics against segmented target population or against full target population, comprising of big data with structured and unstructured data formats.
- IT has to leverage to automated tools, solution frameworks and accelerators to automate the predictive analytical processes.
- IT's role also extends in provisioning highly secured data warehouse and predictive analytical platforms that can ensure full compliance to data privacy and associated regulatory requirements, while dealing with both internal and external data sources (Big Data sources).
- IT can play role in reducing the overall TCO of BI and predictive analytical platform, by providing innovative solutions including emerging cloud based predictive solutions.
- IT's role is critical in facilitating business continuity and scalability of the BI/Predictive analytical platforms, that are critical to business operations in insurance.
- IT can be instrumental in promoting and familiarizing the data mining and predictive analytical practices across the organization, by deploying user friendly and business friendly predictive analytical tools and solutions.
- In the future, predictive analytics moves further towards handling more and more complexity in the models, and IT can play a major role in facilitating the predictive analytics, using models with higher degree of complexity.

8. Conclusions

Predictive analytics can help the insurers to accurately predict the policy, claim, subrogation, customer profile risk scoring models, where such scoring models, can be used to determine optimized policy premiums (policy rates) and to offer faster claim handling and settlement processes. Predictive analytics for claim subrogation and claim risk profiling can reduce the overall claim handling operational costs, thus increasing the overall profitability of the operations. Predictive analytics will also help the insurers to accurately segment their customers, based on their affinities and preferences, where such differentiated segments can be leveraged, to offer tailor made services to the customers, resulting in higher customer satisfaction and higher degree of customer retention. Such precise customer segmentation can be effectively be leveraged for marketing and cross selling purposes. Claim fraud is a major issue to the insurance industry and predictive analytics can help to minimize such frauds across the industry.

Predictive analytics can also provide better visibility, to the risk management units, thus enabling them to adopt a very effective risk management and compliance framework in the organization.

Predictive analytics against big data, originating from external sources, can provide very useful and actionable insight to the organizations.

Currently, multiple BI analytics vendors offer complete set of tools and packages that can support the full cycle of predictive analytical processes, which work against medium as well as big data sources, consisting of both structured and unstructured data. Such tools and packages simplify the complexities associated with the predictive analytical modeling solutions, making it possible to adopt practically feasible predictive analytical solutions. MPP, hardware based and in-database based ETL/BI and predictive analytical solutions, are increasingly being used to perform predicative analytics against large big data sets.

Large scale availability of customer activity data, through unstructured external sources, such as web click streams, social media sites, internet blogs and radio frequency activity channels, have opened new dimensions in predictive modeling space, providing ample of new opportunities to understand their customer preferences and to discover vast amount of new knowledge, that can benefit both the business and its customers. Insurance organizations can immensely benefit by deploying suitable actionable strategies that can leverage to the knowledge provided by predictive analytics. In future, predictive analytics will continue to address higher degree of complexity in the real world, and will continue to improve the accuracy of models and the associated predicted outcomes.

About the Author



Ashok Goudar

Associate Vice President & Senior Architect, Europe

Ashok Goudar is an experienced professional operating in the consultancy services industry for the last 17+ years. He has an in depth experience in business technology/architecture consultancy and program/engagement management services. Prior to joining MphasiS in 2008, Ashok Goudar had worked in EMEA, NA and India regions with two other (TCS, Infosys) consulting service organizations in the capacity of business technology architectural consultancy and IT management roles. His experience includes enterprise solutions (BPM, SOA, Cloud, CRM, SCM, BI/KW and Web) covering Applications, Infrastructure and IT service delivery space applied to Insurance, Finance, Retail and Telco domains.



ABOUTMPHASIS.

Mphasis is a \$1 billion global service provider, delivering technology based solutions to clients across the world. With over 41,000 people, Mphasis services clients in Banking and Capital Markets, Insurance, Manufacturing, Communications, Media & Entertainment, Healthcare & Life Sciences, Transportation & Logistics, Retail & Consumer Packaged Goods, Energy & Utilities, and Governments around the world. Our competency lies in our ability to offer integrated service offerings in Applications, Infrastructure Services, and Business Process Outsourcing. To know more about Mphasis, log on to www.mphasis.com

For more information, contact: sales@mphasis.com.

USA: 460 Park Avenue South, Suite #1101, New York, NY 10016, USA
Tel.: +1 212 686 6655, Fax: +1 212 686 2422

UK: 88 Wood Street, London EC2V 7RS, UK
Tel.: +44 20 85281000, Fax: +44 20 85281001

AUSTRALIA: 9 Norberry Terrace, 177-199 Pacific Hwy, North Sydney, 2060, Australia
Tel.: +61 2 99542222, Fax: +61 2 99558112

INDIA: Bagmane Technology Park, Byrasandra Village, C.V. Raman Nagar, Bangalore 560 093, India
Tel.: +91 80 4004 0404, Fax: +91 80 4004 9999



1011



MPHASIS
an HP company