Beyond Big Data: How to Turn Data into Actionable Insight

By Wade Bontrager, CEO, EagleEye Analytics
Is big data the answer?
We spend an awful lot of time talking about data in insurance, especially big data. Big data is the thing. Big data is huge. Big data is the best thing since the last best thing we had.

When something has so much potential and so much attention, it’s easy to get hung up on the hype. But how is more data going to make insurance organizations successful? Is the data itself going to get us where we need to be? Probably not.

The answer to what?
At a recent industry meeting we talked for two days about all the new data now available, what is the right data to collect and use, and how to get the data right. We never got to the agenda items covering the right way to analyze data and apply it to business problems. Or even what those business problems are.

I spend a lot of time talking to insurance executives, and it seems most of their strategic problems boil down to one of these two themes:

- How to maximize profit without sacrificing growth
- How to maximize growth without jeopardizing profit

We know the dials to turn to increase growth – sales channels, marketing efforts. We know the dials to turn to increase profit – expenses, retention. What we have trouble with is the balancing act, making sure turning up one doesn’t push the other down. The result can be underpriced policies that kill loss ratios and overpriced policies that walk away.

I don’t believe big data is the answer to everything, but with the right tools and approach, predictive insights gained by insurance data mining can be the answer to the growth/profitability dilemma.
Making sense of rate inaccuracies

Here's a good example of the growth/profitability dilemma in real life. Even in the personal auto insurance space – the line of business most advanced at using data and analytics – pricing inaccuracy can be glaringly obvious.

Almost everyone has seen the popular series of video ads with onscreen comparisons showing the advertiser’s rates far lower than the other brands. This Youtube commenter expresses what all consumers are probably thinking when they see ads like these:

“Trwent” is probably right. The rates shown in these ads and in online car insurance quoting tools can’t all be accurate for each customer.

When we tested the accuracy of some actual auto quotes with a set of rates from multiple carriers on three policies (a subset of the data is shown on the next page), the results were stunning.

For Policy 1, rates range from $349 to $802 – a difference of $453. Policy 2 has a similar range. For Policy 3, the range of quotes is staggering – from $764 to $3,582.
When rate disparities are this large, it's clear that some of the rates aren't accurate, even though they are set by using essentially the same tools with similar methodologies. Each brand would probably tell you their rates are accurate, maybe off by ±5%. But this is not 5%.

If the winning brand in the Policy 3 quoting war is the lowest price brand, then that brand will most likely be the biggest loser when it comes to the policy's ultimate loss ratio.

And if auto insurance carriers using not only massive volumes of data but also incorporating leading-edge data like telematics can't get it right, then how can anyone else? Why isn't all this new data improving their pricing accuracy? Why can't they do a better job of predicting the future profitability of new business?

The right tool to do the job right
Every home improvement fanatic knows you have to have the right tools to do a job right. The same applies to mining data for predictive insight – start with the right tools.

The insurance industry isn't exactly notorious for early adoption of new technology. We've been very dependent on spreadsheets, powerful as they are in the right hands, for more than two decades. But spreadsheets are insufficient for extracting predictive insights from data.

On the other hand, the pace of adoption for Business Intelligence (BI) and Data Visualization tools has increased dramatically since the data explosion began. But those tools are best at allowing us to slice and dice views of what happened in the past, and are, like spreadsheets, very limited in their predictive power.

Popular tools like Generalized Linear Modeling and Min-Bias techniques help determine the right rating factors without double-or triple-charging a single risk, for example a 22-year-old single male driver with bad credit and multiple accidents. The GLM spreads the risk among the different

<table>
<thead>
<tr>
<th>Brand</th>
<th>Policy 1</th>
<th>Policy 2</th>
<th>Policy 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$596</td>
<td>$0</td>
<td>$2,015</td>
</tr>
<tr>
<td>2</td>
<td>$373</td>
<td>$292</td>
<td>$1,909</td>
</tr>
<tr>
<td>3</td>
<td>$463</td>
<td>$292</td>
<td>$2,278</td>
</tr>
<tr>
<td>4</td>
<td>$777</td>
<td>$599</td>
<td>$2,713</td>
</tr>
<tr>
<td>5</td>
<td>$604</td>
<td>$560</td>
<td>$0</td>
</tr>
<tr>
<td>6</td>
<td>$381</td>
<td>$253</td>
<td>$1,022</td>
</tr>
<tr>
<td>7</td>
<td>$398</td>
<td>$229</td>
<td>$764</td>
</tr>
<tr>
<td>8</td>
<td>$802</td>
<td>$360</td>
<td>$2,135</td>
</tr>
<tr>
<td>9</td>
<td>$349</td>
<td>$306</td>
<td>$3,582</td>
</tr>
<tr>
<td>10</td>
<td>$479</td>
<td>$306</td>
<td>$0</td>
</tr>
<tr>
<td>11</td>
<td>$676</td>
<td>$326</td>
<td>$0</td>
</tr>
<tr>
<td>12</td>
<td>$555</td>
<td>$416</td>
<td>$952</td>
</tr>
</tbody>
</table>
factors. But because the GLM can't identify interactions in the data, 22-year-old single males with bad credit and multiple accidents get treated as a distinct group. The power of that variable combination is then diluted instead of being spread out.

Perhaps it's time we looked to market leaders in other industries for a new idea.

**Machine learning drives the best predictive modeling tools**

Machine learning selects your next streaming movie or song. Hedge funds use it for automated trading. It's the decision engine in driverless vehicles. It's faster, more accurate, and more predictive than other analysis tools. Machine learning is automated modeling that learns as it goes, applying algorithms in an iterative fashion. It searches through hundreds and thousands of data interactions, creating and testing model after model to ultimately produce the most reliable predictive model possible from the data.

Machine learning isn't a new concept, but new applications of machine learning are changing the fortunes of nearly every industry. Need proof? Watch Netflix.

Consider how machine learning works in online movie streaming. We don't know Netflix’s algorithms, but it's easy to observe what happens. As you watch movies, Netflix tracks what you watch, what you abandon, what you put in your queue and whether you watch it. It uses that information to present viewing suggestions for you – and for other people with characteristics similar to yours.

Viewing suggestions aren’t the only use of machine learning for a company like Netflix. Companies in every industry can use machine learning to identify the attributes of customers who reliably pay their bills (or don't), and target their marketing accordingly. The possibilities are endless.

**Machine learning applied to insurance**

Machine learning works the same way for insurance. Using data on existing policies and claims, machine learning algorithms find the interactions of variables associated with a stated result, such as retention or loss ratio.

Machine learning is equally effective for large and smaller insurers. While it efficiently handles vast volumes of data, it can also produce highly credible models from smaller data sets.
Because machine learning is automated, with massive computing power, it can model huge quantities of data very quickly. In any data set, hundreds and even thousands of possible data interactions exist, and machine learning searches through all of them. It learns from each repetition, or iteration, refining until a final model is produced to represent the most predictive signal in the data.

Could a person try to do this? Yes, and they do, to some extent, with Generalized Linear Modeling. But the approach is largely trial and error, and depends heavily on human intuition pointing the model in the right direction. Machine learning simply uses computers to do what computers are good at doing: Checking out the many possible combinations. Machine learning also lets humans do what they are good at doing: Interpreting what the model is and isn't telling you and deciding what to do next.

Let's look at some examples where machine learning found predictive insights that were immediately actionable.
Turning predictive insight into actionable insight

Example 1

**Analysis goal:** Who are our most profitable customers? What are the characteristics we should look for in new business?

**Predictive Insight:** In a data set from a private passenger auto carrier with a total portfolio loss ratio of 71%, even without machine learning it was easy to see that policies where a safe driver discount was applied had a loss ratio of 87%.

<table>
<thead>
<tr>
<th>Safe driver discount</th>
<th>Loss ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>87%</td>
</tr>
<tr>
<td>No</td>
<td>59%</td>
</tr>
</tbody>
</table>

Considering those without the discount had a far lower loss ratio of 59%, it seems clear that the carrier was over-discounting.

Using machine learning to analyze the variable interactions, many more refined segments emerged. One segment stood out for its extremely low loss ratio compared to the 71% portfolio loss ratio.

A segment of policies with the safe driver discount that also had 10 years or more driving experience, but did NOT have a passive restraint discount, actually had a surprisingly low 35% loss ratio. The segment would have been difficult to find or test using traditional methods, and was a credible segment that was statistically valid.

**Action:** If the carrier reduced everyone’s safe driver discount, many drivers who have the discount would be penalized and may shop for a new policy.

Instead, the carrier has implemented a loss ratio scoring process to target this valuable segment, and others with characteristics that are predictive of low loss ratios, to grow new business and improve retention.
Example 2

Analysis goal: Identify the predictors of the worst loss ratios so potential “loss ratio killers” can be identified and avoided.

Predictive Insight: In another example from the same data set, policies with one or more male driver had a higher-than-average loss ratio of 75%.

<table>
<thead>
<tr>
<th>Male drivers</th>
<th>1+</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss ratio</td>
<td>75%</td>
<td>66%</td>
</tr>
</tbody>
</table>

Policies with drivers with less than two years’ experience had an even higher loss ratio of 90%.

<table>
<thead>
<tr>
<th>Driving tenure (yrs)</th>
<th>&lt;2</th>
<th>2+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss ratio</td>
<td>90%</td>
<td>63%</td>
</tr>
</tbody>
</table>

Neither of these is surprising, and a logical next step would be to look at the different combinations: Less than 2 years with no male drivers vs. <2 years with male drivers, 2+ years with no male drivers vs. 2+ years with male drivers, etc.

Machine learning takes that next step, and many more, by looking at combinations of all variables included in the data set.

Analyzing not only these two variables (tenure and # male drivers) but also adding three more variables yields a segment with a 150% loss ratio:
- One or more male drivers
- Less than two years’ driving experience
- One or more unmarried driver
- Minimum driver age less than 50
- Vehicle age six or more years

Action: Using the model to identify high loss ratio policies at the point of quote could allow the insured to avoid unprofitable business.
Example 3

**Analysis goal:** Create a model to identify large losses as early as possible after a claim is reported.

**Predictive Insight:** Let’s look at one more example, this time in liability claims. Claims organizations everywhere have their checklists of the flags that most frequently indicate large loss potential. These checklists have developed over many years of experience.

This carrier set a threshold of $100,000 as defining a large loss. Looking back at the data after claims closed, we could see that at just 10 days after the claim was reported the carrier’s method had identified 12% of the large losses.

Using machine learning on the same set of claims, however, identified 65% of the large losses using the information available at the 10-day mark.

**Action:** Every week, the insurer runs all open claims through an early-warning claims severity model. Claims flagged go to adjusters with the appropriate skill level and authority to apply claims management resources. This approach ultimately saves 5-10% on average for large losses – a savings that can add up quickly.

**Predicting the future of insurance**

It doesn’t take machine learning to see that every day it becomes clearer that insurers who do not embrace analytics to mine predictive insight from data – and then do something about it – are going to be left behind by competitors who do. Some are way ahead on this journey, and others haven’t gotten started.

Many are stalled on their data analytics journey by questions about data quality, or distracted by the potential of big data. While those factors are definitely barriers to extracting predictive analytics using traditional tools, with machine learning, you can still get started now, with existing data, and find actionable insight that could make the difference in the growth vs. profitability balancing act.
About EagleEye Analytics

EagleEye Analytics offers the most advanced predictive analytics software and consulting services available to the property and casualty insurance industry today. The quantifiable results from the company’s unique machine learning capabilities enable insurance carriers to make tactical and strategic decisions that help drive profitable growth and reduce operating expenses.

EagleEye’s customers have achieved quantifiable and meaningful results by applying modern predictive analytics across the most important insurance business processes – product marketing and distribution, pricing and underwriting, and managing claims. The firm’s solutions have proven to give the best predictive results in over 200 engagements in the United States, Europe, Canada and Australia.

EagleEye is headquartered in Columbia, South Carolina and has a subsidiary in London, UK. A privately held company, EagleEye Analytics is backed by FirstMark Capital in New York. For more information, visit eeanalytics.com

Proud to partner with:

Visit eeanalytics.com today to find out more, or give us a call at +1.803.758.2536