AnalyticsIQ

How to Create Marketing Models That Surpass Your Goals

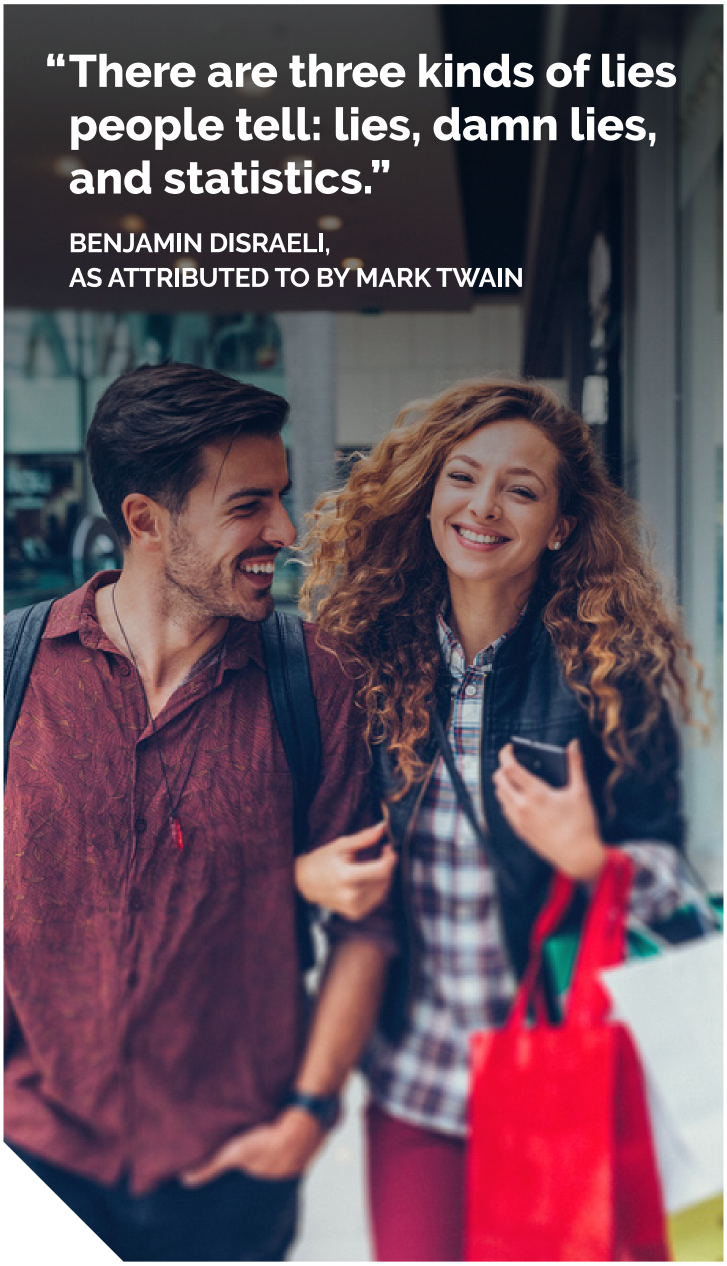
And Avoid Common Modeling Mistakes That Turn Good Statistics Bad

Modeling Is Here To Stay –So We Better Get It Right

Since the 1950's, when credit card companies first began using credit models to predict which consumers would be willing to repay their debts as agreed, statistics have been an important business tool in improving the bottom line for all types of businesses. In the 1970's and 1980's, banks adopted the use of models and "cut-off scores" to approve loans in a fast, efficient, accurate, and (just as importantly) non-biased way. The 1990's saw these methods adapted by marketing firms and credit unions, and today, "analytics" and "logistics" are common industry buzzwords that seemingly everyone is trying to drop into casual conversation.

Statistical models have proven to be reliable and profitable over time, so their adaptation in the business world is no surprise. What's also unsurprising is the rise in the number of statistical modeling failures: models that should never have been built in the first place, that predict completely wrong results, that lose massive amounts of money, and that cause careers to end abruptly.

Most of these failures should/could have been caught early and corrected, but due to a series of poor decisions, they were allowed to fester and destroy an organization from within. The great majority of these failures can be split into three distinct categories: Poor Model Design, Bad Data, and Incorrect Modeling Techniques.



AnalyticsIQ

Poor Modeling Design Pitfalls To Avoid

By far the biggest cause of statistical modeling failures comes before the first dataset is opened, and before the first computer is turned on. Poor modeling design comes from upper management, who either define the problem they're trying to solve incorrectly or don't have a full grasp of their actual goals for the modeling project.

Predicting the Past

It's hard for companies (especially large ones) to always be at the forefront of change in their industry. Change is a constant in business, and identifying the changing trends and pointing the company in the right direction each time market forces are at work is nearly impossible. For many companies, it's easier to stick with what's worked in the past and figure out how to adjust strategies "later" (i.e.-when something goes wrong and there's no choice but to adjust).

A medium sized bank may have been making auto loans quite successfully for many years. Due to competition in the market, they've decided to expand into motorcycle loans and boat loans. Since they don't have any data on these new programs yet, they decide to update their auto models and use them for all of their "motorized" loans. Sure, motorcycles and boats are different, but they can always "adjust" their cut-offs for those later.

A retail company that has been located in the NE United States has opened a new market in California. Since they have no sales data in CA yet, they decide to use their current New England models to set strategies for California "for now".

A reseller in a new and very volatile industry wants a model to predict which of several nationwide carriers would be the "best fit" for individual consumers. However, they insist that all results should maintain the current market share that each of the carriers has, rather than reflect the predicted market share that these carriers will have in the near future.

Each of these companies are trying to be conservative in their approach, but in each case, relying on what they've done in the past, and expecting it to hold true in the future, is a costly mistake. The bank is counting on the fact that boat and motorcycle borrowers perform a lot like car buyers (they don't). The retailer believes that their experience selling, let's say, snow shovels, in Vermont in January will carry over to California (it definitely won't!). The reseller believes that today's market shares will be maintained going forward, despite all evidence to the contrary. In these cases, the past and the future are radically different, and trying to force their new business ventures into their "business as usual" strategies won't work.

Narrowcasting

Closely related to Predicting the Past, Narrowcasting is when a business becomes so focused on what's worked in the past, that they find themselves with an ever-shrinking sliver of data on which to develop their strategies.

When a company first decides to use a marketing model to predict which people to mail, the most common solution is to look at their current customers vs. a random sample of the general population. This type of "cloning model" allows the company to immediately identify people who "look" like their current customers and focus their mailing efforts on them.

Once this works for 6-12 months, the company should have enough data to build a formal marketing model, in which "goods" are those consumers who became customers as a result of being mailed, and "bads" are those mailed individuals who weren't interested. Using this new model, maybe only the top scoring 30% of the people are mailed from now on.

A year or two later, the company decides that it's time to update their successful marketing model by building a new one. Since there's no reason to fix what isn't broken, they use the same strategy as their previous model; "goods" are people who were mailed and responded, and "bads" are people who were mailed but did not respond. The problem is that since everyone who was mailed ("goods" and "bads") were all in the top 30% of the scores from the previous model, our new modeling population is only that fraction of the population. We're assuming that the other 70% of the population is still not interested in the company's product, and that nothing in the market has changed at all.

The new model, then, is focused on this 30% of the population only, and the results of this model lead the company to decide to mail only the top 30% of THAT model.
As you can see, we now have a mailing population of 30% of 30% of the overall market.

The next iteration of models will do the same, further narrowing the focus of the company. Eventually, the company could develop a fantastic, statistically pure model, but they have so few people in their targeted universe that they can't afford to keep the doors open.

In order to avoid Narrowcasting, companies need to continually refresh their data by mailing a small number of people who score in the lower portions of a score, for control purposes. Any model that retreats from the overall population of the company's business footprint runs the risk of Narrowcasting.

Moving the Goalposts

A modeling project may take several months to complete, from the first sample design meeting, when management gets together and decides on what the goal of the project should be, to final acceptance of the model, when results are reviewed, approved, and the model is put into production. It's important to document the sample design meeting, because goals can change over time, and having a written record of decisions made at inception can explain why the final model looks the way it does..

A bank may decide to build a marketing model to predict which checking account customers would be most likely to open another account with the bank. This second account may be savings, money market, auto loan, investment account, IRA, etc. By building a series of models, they can predict that an advertising insert in Person A's monthly statement about savings accounts is more likely to lead to Person A opening a savings account with them, whereas an advertising insert in Person B's monthly statement about IRAs would work better for that customer. These models are part of a "next-best-product" solution.

Once the models are built, someone in upper management states, "No, no, no, what we REALLY want is to increase savings accounts. We want to mail everyone advertising inserts about savings accounts and find some way to boost response rates for them." That's a completely different modeling problem, and not the one that was originally designed. Without a written record specifically stating the goals of the project, this kind of last minute change can easily happen. In most cases, management doesn't even realize that the goal has changed. They just "remember" the sample design meeting differently than the analysts who built the models, so changes will need to be made. Sometimes, the project can be salvaged, but usually, it's a case of starting over from scratch.

AnalyticsIQ

Wrong Performance Definitions

When a bank is developing risk models, it's usually easy to identify a "good" as someone who's always paid their bills as agreed, and a "bad" as charge-offs, bankruptcies, and foreclosures. Usually, 90+ days delinquent will be thrown in with the "bads", as well as 60 days past due (maybe). How about 30 days past due? Is that a "good", a "bad", or an "indeterminate"? That's up to the bank and the specific model they're working on.

For marketing models, the definition of what's a "good" and what's a "bad" may be a little murkier.

> Let's say an insurance company sends out a mailing advertising their life insurance product. Responders are "good" and non-responders are "bad". However, what if the life insurance mailing is an offer for more information about this valuable product? The non-responders are still "bads". The responders are then sent an expensive, glossy 4-color mailing detailing the benefits associated with the life insurance product. Some of those responders end up buying the product, and are "goods".
>
> What about the responders who did not end up buying the product? They were initially interested, so the cost of the original mailing was worth it to the life insurance company. But since they didn't end up buying the product, the company wasted a lot of money on that glossy booklet that they sent them.
>
> Are these people still "good"? In reality, they cost the insurance company a lot of money, more than the non-responders ("bads"). How do we define these "responders-but-not-customers"?

The answer to this question is the dreaded, "It depends". How to handle these not-quite-goods/not-quite-bads depends on the company, the product, past experience, and goals for the project. Of all the potential failures that Poor Modeling Design can cause, this one is the toughest to control, as this is a problem where even experienced managers and statisticians can reasonably argue either side. For this reason, Wrong Performance Definitions have probably killed more projects and created more incorrect statistics than all of the other above problems combined.

Modeling Success Checklist

Want to avoid common pitfalls in modeling? Avoid these situations.

1. PREDICTING THE PAST

As your business and product offerings evolve, be sure to not rely too heavily on past consumer behaviors and model performance. They may not be directly applicable to your new ventures.

2. NARROWCASTING

Narrowcasting may provide you with a high performing model, but there is a good chance you will lack the scale to reach your business goals. Refresh your data often and make sure your audience size is in proportion to your business objectives.

3. MOVING THE GOALPOSTS

With many key stakeholders often involved in the development of a marketing campaign or modeling initiative, too many organizations change their goals at the last moment. Ensure everyone is on the same page (and the goal is in writing!) before beginning your modeling work.

4. WRONG PERFORMANCE DEFINITIONS

Make sure your team is aligned on what constitutes a good or bad consumer response so you have a clear understanding of how to evaluate the model's effectiveness, and how to use (or not use certain data) the next time around.

Bad Data Can Have A Big Impact

Once a project has been defined, it's up to the company to come up with the data to be used for modeling. This data may come from a wide range of sources, and may include individual, household, neighborhood, zip code, county, MSA, or state data. It may include a mix of client-specific data as well as data from other sources, purchased for modeling purposes. It's very important that this data matches the goals of the project, though, in order to have results that actually mean something. The two problems to look out for here are Wrong Data and Skewed/Biased Data.

A company sells their product in northeastern states and specifically mails offers to households making $100,000+/year. They've decided to expand their market by mailing to households making $50,000+/year and decide to build a new model.

All of their "goods" are people who make $100,000+/year and responded to their previous offers. The company doesn't keep track of their non-responders ("bads"), so they decide to build a cloning model by taking a random sample of the population. Since they want to start mailing to people who make $50,000+/year, they pull a random sample of the US population making $50,000+/year as their new "bad" sample, and build their model. The results look fantastic! (There are several ways to measure how well a model can separate "goods" from "bads" – KS, area-under-the curve, etc. For our purposes, we'll go with KS.) The KS report shows that their new model does great in predicting who will become a customer and who won't. The company can't wait to put this model into production so they can start reaping the profits from their new strategy.

There are a couple of problems here. The main problem in this situation is that the client is using the wrong data for the "bads". Of course the model "works"; "goods" and '"bads" look nothing like each other! Because the random data has people who make half of what the "goods" do, the model SHOULD look great. In fact, it's probable that the entire model could be defined by a single variable: Income.

In reality, though, results will be much worse. The model in its current form is going to severely penalize lower income (close to $50,000/year) people and greatly reward people making $100,000+/year. And why wouldn't it? The only people in the sample who made less than $100,000/year are "bads", and ALL of the "goods" are $100,000+/year earners. Obviously, low income equals low (or no) response. The model will never accurately reflect the market. Because of this, management may quickly come to the conclusion that the lower income initiative is a failure and they need to stick to just the well-heeled individuals.

The correct way to develop this model would have been to match the "bads" with the "goods". In other words, if all the "goods" have incomes $100,000+/year, then the sample of "bads" should have the same criteria. In essence, they would be building a model to predict how well their current strategy of $100,000+/year works. Once that model is complete, they should use it to slowly lower the income requirement, first to $90,000, then $80,000, and finally down to the desired $50,000/year threshold. As the income requirement is lowered, they should carefully watch scores of the new people who are responding, and adjust cut-offs accordingly. This isn't an overnight change, but takes months to do. Once the income threshold has been lowered to $50,000/year and several months have gone by (where there are enough lower income "goods" for analysis), a new model can be built, looking at "goods" and "bads" of $50,000+/year.

Unlike the situations discussed in previously in this whitepaper, the company in this case is starting the project out knowing that they are going to be facing an unknown result with a new strategy, and are preparing to face it head-on. They can use their current strategy as a starting point. Rather than blindly setting cut-offs that have worked in the past and sticking with them, they begin by constant experimentation of adjusting cut-offs, offerings, and new client acquisition strategies. The current models are used as a starting point, not as the final judgment. New data is collected continually, with the idea of building an updated model (incorporating the new client base) as soon as possible. Many times, there may be several iterations of the new model, as more data comes in and new experiences need to be included.

The second problem presented in this example is one of skewed, or biased data. Note that the company sells their product only in the northeastern United States. If they pull the random sample from the entire country rather than from the northeastern states, there is a possibility of having a biased model. In this case, the results could be skewed because "bads" include people from Hawaii, Alaska, CA, etc., whereas "goods" only contain people from the Boston-New York-Philadelphia area. In many cases, this may not matter, especially if the product in question is a mainstream one that people around the country typically purchase. If the product is unique to the northeast (a regionally famous BBQ sauce, snow shovels, Red Sox memorabilia, etc.), then there could be a bias problem, and the sample of "bads" should further be limited to people in northeastern states who make $100,000+/year. In order to test for bias, run frequencies and means analysis on "goods" and "bads" prior to modeling to see if area of the country may skew the results.

There are other types of potential data biases as well. Many times, common sense, experience and logic will need to come into play.

The best way to avoid Bad Data issues is to think things through in the early stage of the statistical process. Asking questions for how to "pull" the sample of "bads" needs to be discussed and agreed upon by all parties before analysis begins.

AnalyticsIQ

Improper Techniques Can Lead To Poor Results

Statistics can go bad once the modeling process has commenced, especially when there's confusion over which statistical techniques should be implemented for the client's solution.

Using the Wrong Tools

Using the wrong tools isn't just a problem in the construction or plumbing industry. A company needs a solution to their marketing problem, but does the proper solution involve a multivariate regression (linear or logistic), decision tree analysis, segmentation, clustering, simple criteria, or something more exotic? Possibly, a combination of several of the above choices?

> Occasionally, a project will go off the rails when terminology gets confused. A marketing manager may tell his analysts that he needs to break their customer base into 5-6 clusters. The statisticians go to work creating these clusters, only to find that the manager only wanted 5-6 segments, based specifically on age and gender. When a statistician has spent the time necessary to build accurate and statistically significant clusters, and his manager asks, "Which group has the females aged 18-35?", his day is ruined.
>
> Clustering is a type of segmentation, a very specific type, but "segmentation" in marketing usually refers to creating mutually exclusive groups of people (often demographically different). Clustering and "segmentation" can be viewed as two totally different techniques because of this, but many people use these terms interchangeably.

This problem is closely related to another reason for statistical failure; the idea that Complicated = Better. Many times, a statistician, especially one recently out of school, my feel that the more complicated the solution, the better the results. This is almost never the case in real life. In fact, it's generally the opposite, in that complicated solutions tend to lead to more errors/mistakes, which can be buried by the sheer confusion of what exactly all these moving parts are actually supposed to do. Some statisticians are seeking statistical "purity", where any real-world projects must meet all classroom criteria of statistical "fit". Others believe in "kitchen sink statistics", where the more techniques you throw at a problem, the better the eventual solution.

> As part of an academic exercise, I was once part of a team trying to predict which charitable donors who had since lapsed (quit donating to the charity in question) would be most willing to start donating again if prodded, and how much would they be willing to donate. Our solution involved the following: building a "parent model on returning donors vs. non-returning donors, breaking the score into 20 even group, calculating the average dollars per group, building a new model on this average dollar variable, building multiple error term models and using them to transform the original model (a nod to time series econometric analysis), and tying everything back to the first returning donor vs. non-returning donor. We also utilized CART decision trees to create new, combination variables to replace some of the basic variables we were given. This list doesn't include the other techniques we tried that didn't seem to work no matter how hard we tried to shoe-horn them in.

AnalyticsIQ

In the end, our results were sent in and were validated on an out-of-time data sample by the judges. Our results were that we did slightly better than the groups who built a handful of basic models, about even with groups that did a lot of statistical rigmarole like us, and a little worse than the groups that ran a series of mainframe servers for 72 hours straight in order to maximize every available dollar through neural networks. In other words, a lot of extra time was spent for not much extra result. Plus, if we had made an error in any step along the way, would it have been possible to isolate that error and fix it? Fortunately, we never had to find out.

Incidentally, a couple of groups went wild and tried some "cutting edge" ideas. Their validation results proved worse than random chance (!)

There's a famous saying called KISS (Keep It Simple, Stupid). This works in statistics as well as other lines of work. If a project needs to move up to the next level of complication, it should be understood by all why that need exists and what this new level is expected to bring to the solution in terms of additional lift in predictive power.

Over-Fitting Data

When building a statistical model, it's always wise to randomly create a development dataset and a validation dataset. The model is built completely on the development dataset. At the end of the project, the validation dataset is scored on the final model. Ideally, the development and validation datasets SHOULD have extremely similar scoring results (KS, area-under-the-curve, etc.). If not, it's likely that the model was over-fit to the development data.

Over-fitting is simply the creation of a model that works great on the data that it was built on, but it's so specific to that development data, that it's unable to work as well on other similar datasets. Maybe there was a quirk in the data regarding a certain variable, and that variable played an inordinately important part in the final model solution. Sometimes, when dealing with character variables rather than continuous, numeric variables, the statistician may have "cherry-picked" extreme values and created variables from those overly specific values, leading to a great development solution but a terrible validation solution.

The validation sample is great at finding immediate problems with over-fitting, but an out-of-time sample will really do wonders. If possible, build the model as before, but then have a newer, more updated sample of data waiting in the wings. For marketing, maybe the model was built and validated on responders and non-responders from January-March. The out-of-time sample could come from responders and non-responders collected in April-June. If the model validates on this sample, it's a good sign it will work well once it's placed into production. Seasonality can be addressed by choosing your modeling sample from various points over the course of the year, with an out-of-time sample pulled similarly.

Ready To Go To The Next Level?

Statistical failures can happen at any phase of any project. However, with attention to detail, and a willingness by all involved parties to agree on project goals and processes throughout the project, they can be greatly minimized. The examples listed above are only a sample of things that can go wrong in the statistical modeling world; there are many other potential failures hiding around every corner. An alert and seasoned statistician (along with management) can root these out before they cause trouble.

Let's Talk

Are you ready to start using sophisticated data to grow your business? Our flexible approach makes it easy. Whether you are looking to test, build custom models, understand lifetime value, or target prospects across channels, AnalyticsIQ can be your partner.

Contact us today at sales@analytics-iq.com.

About The Author

Gregg Weldon, AnalyticsIQ's Chief Data Scientist, has been with AnalyticsIQ from the start, having also been a part of the successful company, Sigma Analytics. He is numbers all day, every day, which extends into one of his favorite hobbies—poker. No casual player, Gregg visits Vegas every year for multiple tournaments including the World Series of Poker Championship. He is Carolina bred, earning his Bachelors from Clemson and his Masters in Economics from the University of South Carolina. Gregg's favorite data-related factoid: You can't make a Straight with a 5 AND a 10, but it's impossible to make a Straight without a 5 OR a 10. If you enjoyed this whitepaper, you can expect this same level of insight, expertise and guidance from the entire AnalyticsIQ team.

AnalyticsIQ

AnalyticsIQ