IT'S NOT YOUR MODEL, IT'S YOUR DATA

LARRY BEVENS
Principal Consultant,
Financial Services Analytics

HIGH-PERFORMANCE MODELING STARTS WITH RIGOROUS DATA CLEANING AND ANALYSIS

The demand on data science to enable profitable, fact-based marketing decisions continues to accelerate. Financial Services marketers now better realize the power machine learning (ML) and artificial intelligence (AI) bring to change the way brands make decisions and optimize customer experiences with pinpoint precision.

It's not hard to envision a future with no human intervention needed during the entire decisioning process. While the reach of ML and AI capabilities continues to expand, it is easy to get caught up in the promise of data science. But it's important to not lose sight of the critical upfront work required to properly prepare data before analyzing it. Without a rigorous planning phase, these exercises may not deliver the expected business outcomes.

> ## A shortcut is the longest distance between two points.
>
> – Charles Issawi

How often do people receive an email or see an ad that seems completely irrelevant? With so much first-, second- and third-party data available, why does it seem like the brands people have interacted with for years still don't

CASE STUDY

Acxiom's data science team recently collaborated with a client to build a relatively straightforward acquisition model that was to be based on converter bookings over a period of time. Somewhere in this data set, the team uncovered a subpopulation driving a significant portion of the conversion rate. A deeper look found that these people were influenced by a new account bonus offer. Had this bias been left unchecked, the client likely would have been at risk of implementing a model skewed toward a propensity to accept a temporary bonus offer rather than a propensity to convert. If the Acxiom team hadn't done the data deep dive, it would have delivered a well-built model that would not perform as expected.

seem to know them? It's likely the data is not the problem; rather, it's all the disconnects among interpretations of what it says, from what brands hear and the stories they choose to tell.

The idea of building algorithms that activate a population optimized around an outcome is fun, especially compared to the time-consuming work required to complete and connect disorganized data. Quite often the curiosity and grunt work required to expose unrepresentative data, biases and outliers is cut short, resulting in delivery of a well-built model that performs below expectations.

Even with the cleanest, best structured data, brands still need to allow for time and depth of data analysis to increase the likelihood of the models' success. They should include as many elements of the customer journey as possible so patterns, anomalies and their impacts on an outcome can be surfaced, understood and addressed.

Completeness and cleanliness of data is important to building effective models. To make your data work for brands, they must gain a full understanding of the story being told. That will ensure it is consistent and true.

CONCLUSION

The preparation and cleansing of data are as important as analytics. Without a rigorous planning phase, models and analytics can be biased by outliers and anomalies. As marketers take on new capabilities like AI and ML, they can't forget the truth that lies in clean source data.