

BUILD A SUCCESSFUL PROPENSITY MODEL TO GROW REVENUE

By Brendan Meany, James Lee, Tad Zhang & Arvid Tchivzhel

February 2019

In this article, we will give a short summary of machine learning and econometrics in context of propensity to subscribe. We will also share results and ongoing use cases of how publishers are leveraging predictive modeling in their acquisition strategy.

As publishers continue to become more sophisticated in their audience development strategies, analyzing propensity to subscribe is essential to find and target the right audiences. There are many methodologies that analysts have used to predict a user's willingness to subscribe, among them: machine learning, lookalike modeling, clustering, and econometrics. These techniques have tradeoffs so must be used appropriately for the kind of question or problem being solved.

The first three techniques listed above can be grouped into the broad category of data mining. This approach is best used to identify trends and classify which users have a high propensity to subscribe.

Econometrics uses similar mathematical principles but is driven by a hypothesis testing framework. This approach is generally used to estimate the nature of the relationship of variables on propensity to subscribe.

Understanding the distinction between machine learning and econometrics allows publishers to leverage the best of both methodologies to understand and target their audiences.

Making the Case for Machine Learning

Along with buzzwords like "artificial intelligence" and "virtual reality", "machine learning" is one of the most popular in the media industry today. Often publishers analyzing propensity will brag about the number of features (another word for inputs) in their propensity models, reminiscent of how car

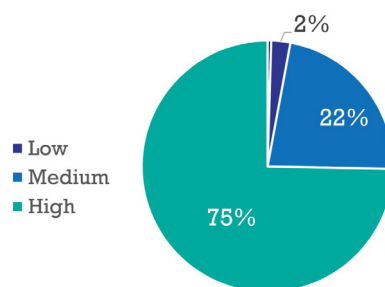
enthusiasts brag about horsepower or torque! While it is important to have a broad array of inputs, the effectiveness of a machine learning model is not determined by feature count alone.

The most important evaluation criteria for a machine learning model is accuracy of out-of-sample predictions.

One simple metric analysts use is the sensitivity of a machine learning model – in other words, did the model correctly predict users who became subscribers. If the model reached close to 100% predictions out-of-sample the model will be successful in identifying future subscribers.

Another important test metric is specificity, which indicates the percentage of false positives. Though this measure can be costly in some industries (such as health care) for subscription conversion, it is less important to consider. The result of having low specificity simply means there are many users who look like they should have subscribed but have simply not yet subscribed. Therefore, focusing on sensitivity and deprioritizing specificity gives publishers a lookalike target audience to begin testing different user experiences and personalization. These users have all the signs of subscribing but perhaps need more relevant calls to action or personalization to capitalize on their propensity.

Actual Conversions by Segment



This chart illustrates how well the machine learning model allocated conversions from the prior month into propensity segments.

75% of the conversions were properly identified as "high propensity" by the model.

22% of conversions were identified as "medium propensity" which can be used to improve the accuracy of subsequent models.

mather: insights

In summary, the major strengths of machine learning are:

- Uncovering trends and patterns that economic theory might not consider
- Higher degree of prediction accuracy than other methods

More areas to consider are:

- Difficulty in explaining “why” a relationship exists
- Overfitting a model due to a biased dataset or insufficient variability

Interpreting the coefficients is the primary use case for an econometric model.

Making the Case for Econometrics

In contrast to machine learning, model specification (picking which variables to include) and interpretation are key components of econometrics. The first step to build an econometric model is to consider what inputs should be in the model based on expected relationships independent of each other. For example, page views and article page views may be expected to both have a positive correlation with propensity but are likely too correlated to both be in the model. So, choosing wisely and testing for any collinearity between the input variables is important.

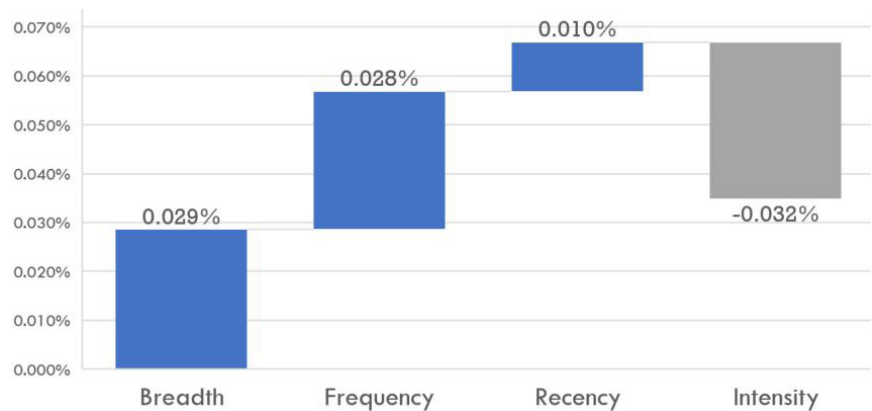
Once the model selection is complete, interpreting the coefficients is the primary use case for an econometric model. If there is a statistically significant relationship in the model, publishers will be able to confidently conclude which variables have the greatest impact on propensity to pay. In addition to statistical significance and magnitude of the coefficient, the direction (negative or positive) will clearly indicate the types of variables publishers should encourage or discourage. There are often intriguing insights on behavior and content that are revealed after applying econometric modeling.

The chart shows the marginal effects of different engagement indicators on propensity to subscribe. Breadth (how many distinct topics a user reads within a session) and frequency of visit (within 28 days) show the greatest positive effect on probability to subscribe.

Interestingly, the intensity (page views and time within a single day) shows a negative correlation with probability to subscribe. Since econometrics estimates the marginal impact of each variable while controlling for all other variables it is likely that page views and time are already correlated with breadth and recency.

This distinction helps publishers focus on improving the right metrics with their engagement and loyalty campaigns.

Marginal Effects by User Engagement



The major strengths of econometrics are:

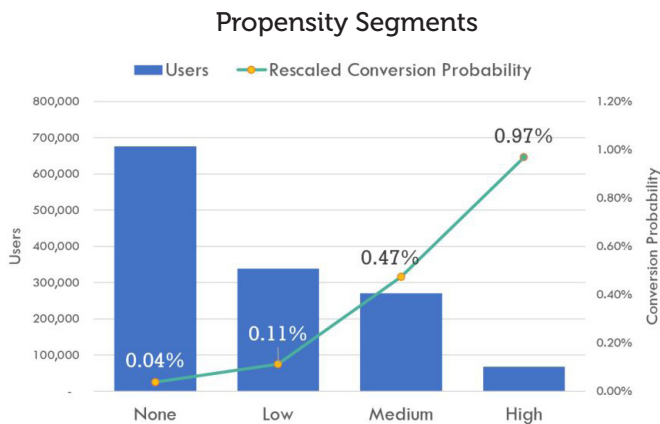
- Explaining directional and statistical significance of propensity to subscribe
- Ability to use insights to make marketing, content, and other tactical decisions

More areas to consider are:

- Relies on analysts for model specification rather than purely data-driven decisions
- May have less flexibility to identify high propensity users in production

Application and Use Cases

After the users are identified and the insights materialize, publishers have a variety of tools and tactics to act on the results. The main application of propensity modeling is usually via the paywall system or email system. Most modern platforms can apply different business rules (such as lower meter settings or personalized messaging) and user journeys targeting each individual user. Mather Economics is supporting A/B testing in multiple markets to measure what types of calls to action and journeys work best to convert all the high propensity audiences, nurture those who show moderate propensity to pay, and engage those with low propensity to pay.

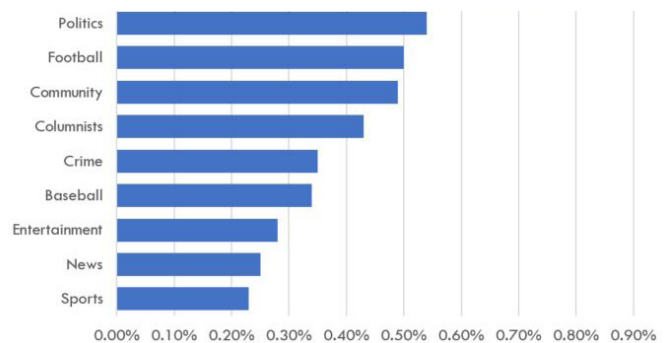


The chart shows performance of the propensity model in production – the publisher can balance ad revenue and user experience to capture consumer revenue from the “high” group while being less aggressive and reducing advertising risk with the “medium” and “low” groups.

In addition to applying the model on user experience through fulfillment systems, publishers can leverage this data in the newsroom. Identifying individual articles read by each group gives writers guidance on what stories resonated most and which ones did not attract the high propensity audience. Editors are also able to understand what topics and content types generate the highest engagement from high propensity audiences. By adjusting the focus of the newsroom, resources can better be allocated to ensure there is enough content produced to improve propensity to subscribe.

Recent studies of propensity and content have shown that audiences value important local stories that journalists seek to cover. Using propensity modeling can spur investment in the types of content that is critical for local markets.

Content by Ratio of Converting Users



The chart above ranks content in an anonymized market by the ratio of page views of users who became subscribers. These users are often categorized into the high propensity audience segment. Econometric modeling also reveals content that shows strong correlation with propensity to subscribe.

There are many materials available that offer a deeper dive into some of the concepts and techniques introduced in this article. The high-level overview of the techniques and applications should give enough to begin asking the right questions and set the stage for success.

By leveraging the strengths of each methodology, publishers can accurately identify users who will pay for content and the factors that are important in explaining why they are willing to pay.

Adjusting marketing tactics, newsroom production, and user experience are some of the ways publishers have utilized predictive analytics. Applying propensity modeling is a core part of the data and analytics required to build a sustainable digital business model.