Using Signal Testing and Predictive Analytics to Improve Marketing ROI

Worthy Technology LLC

worthy.ai

July 17, 2020

Abstract

The rise of machine learning is redistributing control from the advertiser to the algorithm. Savvy advertisers are embracing the loss of control and assuming the role of algorithmic trainer. Facebook agrees that advertisers who have learned to "take their hands off some of the controls can benefit from machine-enabled liquidity—a state in which every dollar can flow to the most valuable impression."[1] In this white paper we:

1. introduce a novel campaign optimization strategy called Signal Testing;
2. review the shift toward algorithmic ad buying which has made Signal Testing increasingly powerful;
3. introduce the Worthy Signal Testing Platform; and
4. review the results from three app advertisers who used the Worthy Signal Testing Platform to develop predictive signals that significantly outperformed their previously dominant signals. Using Worthy:
   (a) Parx reduced its cost per first depositor by 33%
   (b) A top meditation app reduced its cost per paid subscription by 26%
   (c) Abide reduced its cost per paid subscription by 17%

[1] https://www.facebook.com/business/news/insights/boost-liquidity-and-work-smarter-with-machine-learning

Contents

1 Signal Testing

Signal Testing is the continuous testing of the campaign's optimization signal. This optimization signal is the specific result which the advertiser trains the algorithm to optimize toward. This could be the lowest cost per impression, click, install, registration, purchase or literally any result an advertiser can track and feed back to the algorithm. In Facebook's words, "when you make your [Signal] choice for an ad set, you're telling us to get you as many/much of that result as efficiently as possible."[2]

There are an infinite number of potential signals. While an app may appear to only have a handful of meaningful behaviors, an advertiser can fire a signal for any combination of behaviors over any time interval. But advertisers are unlikely to develop performant signals by clustering random combinations of behavior.

Instead, start with the ideal matured customer outcome and work backwards to determine the early behaviors most associated with the outcome. For example, take the case of a subscription app with a 7-day free trial which wants to acquire paying subscribers. Rather than simply optimizing for the Free Trial signal, this app could develop a new signal for the combination of early user behaviors that it predicts is most correlated with becoming a paying subscriber. Or take the case of a real money gaming app, which derives an outsized share of revenue from a small percentage of high value players. Rather than optimizing for the First Time Deposit signal, this gaming app could create a new signal for the set of early user behaviors that it predicts is most correlated with maturing into a high value player. The most successful signal testers use predictive analytics to develop new signals on the basis of which patterns of behaviors predict which outcomes.

Most advertisers initially test a handful of optimization signals, settle on a winner and move on. However, finding a winning signal and never testing new signals is analogous to finding a winning ad and never testing new ad creative.

[2] https://www.facebook.com/business/help/355670007911605?id=561906377587030

Unfortunately testing a new signal is not as technically simple as testing a new ad creative. Signal testing has traditionally required a team of engineers with expertise in both machine learning and event instrumentation; and it has therefore been out of reach for all but the largest and most technically proficient advertisers. However, as the trend toward algorithmic media buying continues ever faster, signal testing will become increasingly important, offering a growing edge to those who have mastered it, and leaving those who haven't behind.

Before we review how Worthy's platform democratizes access to signal testing, it's important to first understand the changes that are making signal testing increasingly powerful.

2 The Rise of FBLearner Flow and Behavioral Optimization

For most of Facebook's history as an ad platform, hyper-targeting was the dominant ad strategy. The savviest Facebook advertisers sliced and diced thousands of combinations of audiences and ad creative in an attempt to discover audience-creative pairings which would yield the highest engagement. This was an effective strategy as Facebook would reward hyper-personalization with significantly lower costs.

However the pendulum began to swing toward a new strategy with the rise of FBLearner Flow. In order to scale algorithmic optimization on their product internally, Facebook "decided to build a brand-new platform, FBLearner Flow, capable of easily reusing algorithms in different products, scaling to run thousands of simultaneous custom experiments, and managing experiments with ease."[3] As Facebook was principally motivated by increasing ad revenue, FBLearning Flow ultimately evolved into a tool for advertisers, allowing them to train Facebook's most powerful algorithms on their unique behavioral data in order to teach the algorithm how to hunt for whatever future behavior they desired.

[3] https://engineering.fb.com/core-data/introducing-fblearner-flow-facebook-s-ai-backbone/

This became clear in Q3 2016 when Facebook launched a new ad product called App Event Optimization. With the advent of app event optimization, Facebook increasingly began segmenting its own audience according to behavioral propensity—that is, segmenting on the basis of what a given user is likely to do in the future. And indeed, within every targetable audience, there is a certain subset that has a propensity to:

(a) take any action from ads

(b) watch videos but not click

(c) click but not take additional action

(d) install but not take additional action

(e) take in-app action but not purchase or subscribe

(f) purchase or subscribe

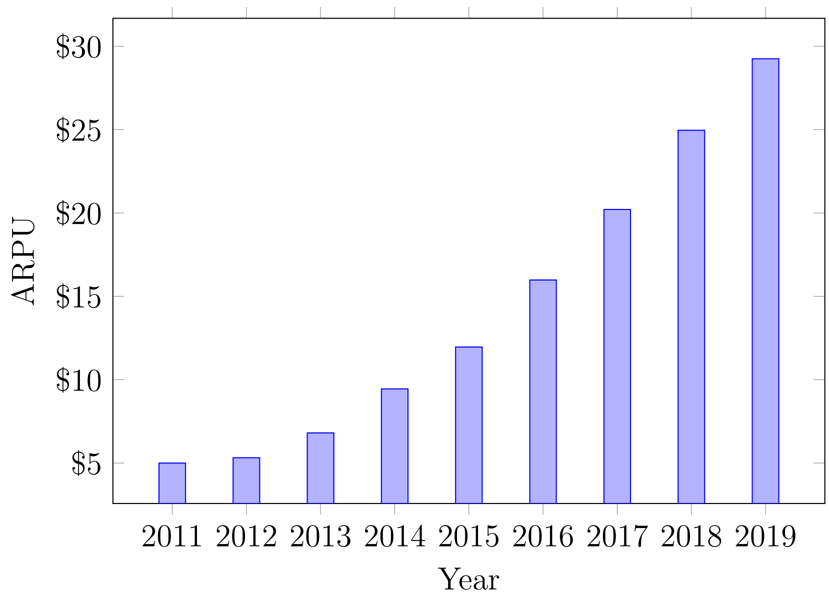
Behavioral propensity is an invisible targeting layer that exists across Facebook and Instagram's entire user base. It is not directly accessible through audience targeting in the same manner as demographic or interest-based targeting. Instead, advertisers must rely on optimization signals to use FBLearner Flow in order to train the algorithm to serve ads to the right subset of their audience based on their desired outcome.

A brand advertiser who is attempting to get the most video views for the lowest cost won't be willing to pay a higher cost for an impression that's likely to convert to sale. On the other hand, a performance marketer who is attempting to maximize directly-attributable return on investment would gladly pay this premium, so long as the return is there.

This dynamic has enabled Facebook to demand the highest possible price for each impression on its suite of platforms by only serving an advertiser's ads to the users most likely to exhibit the advertiser's desired behavior. This is plainly reflected in Facebook's average revenue per user (ARPU) growth: an 83% increase in ARPU between 2016 and 2019.[4]

[4] https://www.statista.com/statistics/234056/facebooks-average-advertising-revenue-per-user

Facebok Average Revenue Per User (ARPU), 2011-2019



The continued ARPU growth was a direct result of being able to charge advertisers higher CPMs, which advertisers gladly paid in exchange for better conversion rates and a higher return on their ad spend.

As advertisers yielded control to the algorithm, explicit audience targeting became a less important factor in campaign performance. Traditionally advertisers would bid for the lowest cost per impression or click and therefore need to use precise targeting to ensure they were acquiring impressions or clicks that would convert. But now advertisers can train the algorithm on a desired outcome and leave audience targeting mostly open, trusting the algorithm to seek out the right users from a near infinite potential audience.

3 Cambridge Analytica and The Great Submergence

The rise of action propensity segmentation and app event optimization on Facebook set the stage for the shift toward algorithmic buying, but it took a global data scandal to truly catalyze the shift.

In early 2018, Facebook had the largest data breach in its history as the world learned that Cambridge Analytica had harvested the personal data of 87 million users, selling the data to political advertisers for psychographic targeting.

Cambridge Analytica purported to have scraped so much data about every user that they knew what each person believed and the best way to change that belief with a specific messaging strategy—that is, whether to appeal to emotion, authority, fear or any particular psychological persuasion lever. In pitching psychographic targeting, Cambridge Analytica pitched the apotheosis of the niche targeting strategy which was originally the dominant ad buying strategy on Facebook.

As a result of the Cambridge Analytica Scandal, Facebook's brand reputation fell more than any other brand between 2018 and 2019, tumbling 43 spots on the Axios Harris Poll 100/[5]

As a result of the backlash, in March 2018, Facebook decided to shut down its Partner Categories program, effectively restricting advertisers from using third party data segments for ad targeting from data brokers like Oracle, BlueKai and Acxiom.[6]

This was a tectonic shift for targeting on Facebook. Before their access was pulled, advertisers could browse a tantalizing array of purchase audiences from all major credit card providers. After paying the data brokers a fee, advertisers could target ads to users who had purchased specific SKUs of products from specific sets of stores. Savvy advertisers could combine these data broker audiences with their first party audiences (e.g matched email lists) or Facebook's free demographic and interest audiences to run hyper-targeted ad creative against an unfathomably large array of niche audiences.

There was never - and probably never again will be - an ad targeting

[5] https://theharrispoll.com/axios-harrispoll-100/

[6] https://techcrunch.com/2018/03/28/facebook-will-cut-off-access-to-third-party-data-for-ad-targeting/

platform as powerful as Facebook's in February 2018. But it was only the advertisers who lost the targeting power.

Because while Facebook removed Partner Categories from advertisers, Facebook didn't stop partnering with the data brokers or using their data. Facebook merely submerged the data, out of reach from advertisers, and into the depths where the algorithms reign.

To be clear, Facebook still enriches user profiles with every imaginable data signal to help its algorithms better predict how users will behave and which ads to serve them. Facebook still has transparency into where you went, what you bought, what you think and who you are.

But now advertisers can only leverage a fraction of Facebook's omniscience for ad targeting. At least not directly, anyway.

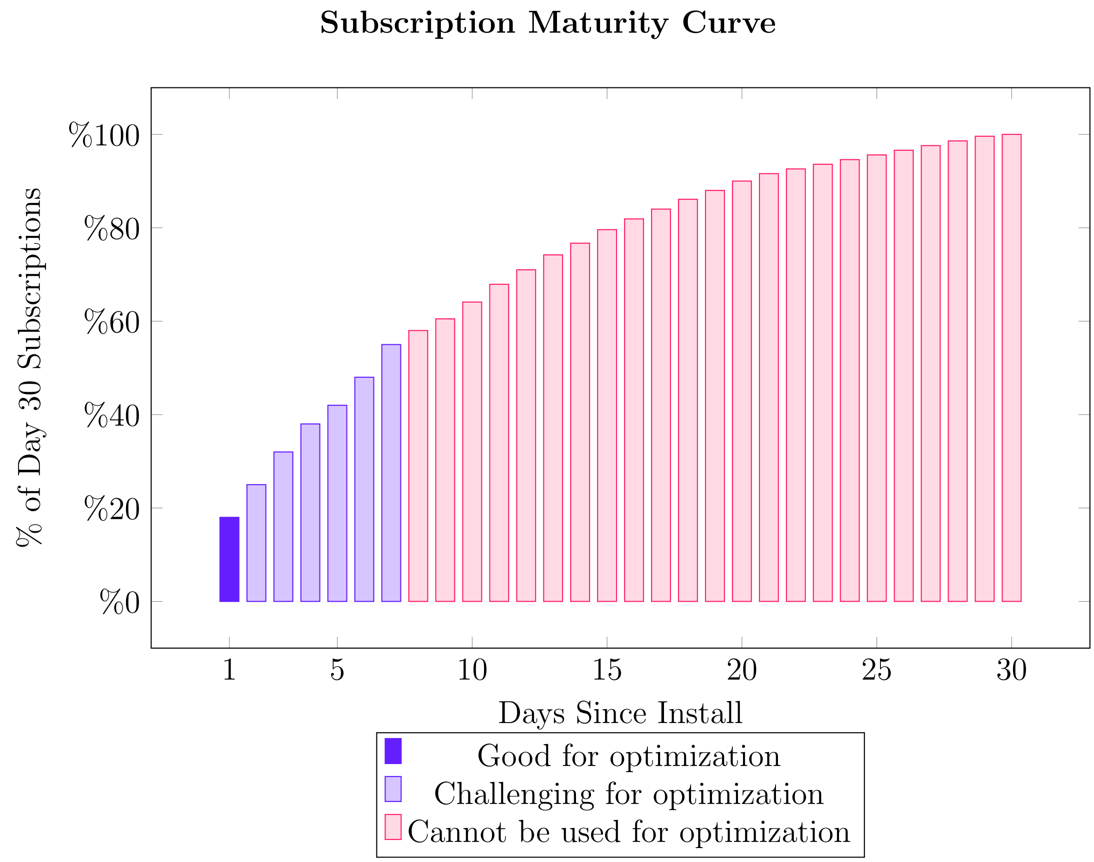
The Cambridge Analytica scandal catalyzed another shift from a world in which Facebook rewarded advertisers for hyper-targeting toward a world in which Facebook rewarded the advertisers who learned to train the algorithm to use the submerged data on their behalf.

4 The Conversion Latency Optimization Challenge

Given that advertisers can now trust the algorithm to optimize for any result they want, why aren't advertisers simply optimizing directly for the lowest cost per lifetime value?

We now encounter the first constraint: the conversion window. Facebook forces advertisers to select a specific conversion window for each result. It's not enough to ask for the lowest cost per lifetime value; advertisers must specify that they want the lowest cost per lifetime value within 1 or 7 days of clicking or viewing an ad. This means that advertisers cannot directly optimize toward any result that occurs outside of 7 days of clicking an ad. And even if the result occurs between 2 and 6 days of clicking an ad, Facebook is ill-equipped to optimize toward it.

Facebook's algorithm is best at optimizing for results that occur immediately after ad exposure. The longer out the behavior is from the ad exposure, the more difficulty the algorithm has optimizing toward the behavior.



This is a genuine challenge for all advertisers whose products take time to mature new users into paying users.

For example, every app with a free 7-day trial is forced to optimize for free trials instead of subscribers because the subscription event occurs just outside of the 7-day optimization window. This means advertisers are training Facebook to seek out the cheapest free trial and not the cheapest subscriber. But the cheapest trial is rarely the cheapest subscriber, so most free trial advertisers waste the majority of their budget on free trials that never convert.

But even if you're not running a 7-day free trial acquisition flow, ask yourself this: why aren't advertisers optimizing directly for lifetime value?

5 Worthy, The Signal Testing Platform

Worthy was developed to equip advertisers with the tools required to power the end-to-end signal testing process. Worthy offers advertisers a platform for powering each step in the signal testing cycle:

1. **Predict Future User Behavior.** The best signals to use for ad optimization predict future user behavior. For example, a completed user profile may be an important predictor of future recurring purchases. A more powerful signal evaluates multiple user characteristics and actions to determine the likelihood of a future behavior.

   Worthy uses machine learning to create a powerful signal capable of predicting future behavior. Each machine learning model combines user characteristics, like device, user actions, like specific user events, and other attributes to form a signal ready for Facebook ad optimization.

2. **Pass Custom Signals To Ad Platforms.** To be able to predict future behavior Worthy processes every user-generated event, equating to more than 10s of millions of events per month for each partner on average. 99% of events incoming to Worthy are accepted into the system within $\frac{1}{4}$ of a second.

   Worthy's cloud data lake provides Worthy with low-latency access to user events, enabling Worthy to generate predictions for new users on-demand. Each custom signal is passed directly to advertising partners, allowing advertisers to optimize their campaigns for custom Worthy signals.

3. **Measure Signal Performance.** Worthy keeps track of all predictions and signals in Worthy's cloud data lake. For each signal Worthy's dashboard allows advertisers to see the accuracy of the signal and a real-time graph of how many times the signal occurs per hour.

   Combined with Facebook ad account data, Worthy also enables advertisers to monitor their split test performance and act when split tests reach statistical significance.

4. **Iterate To Develop Better Signals.** After each split test is complete, the Worthy team works closely with advertising partners to refine custom signals. First, the signal's performance is compared to the status quo looking at costs in the purchase funnel from start to finish. Next, the Worthy team iterates on the signal to create a new signal that may perform better. Finally, a new split test is launched to test the new signal aiming to replace the existing status quo.

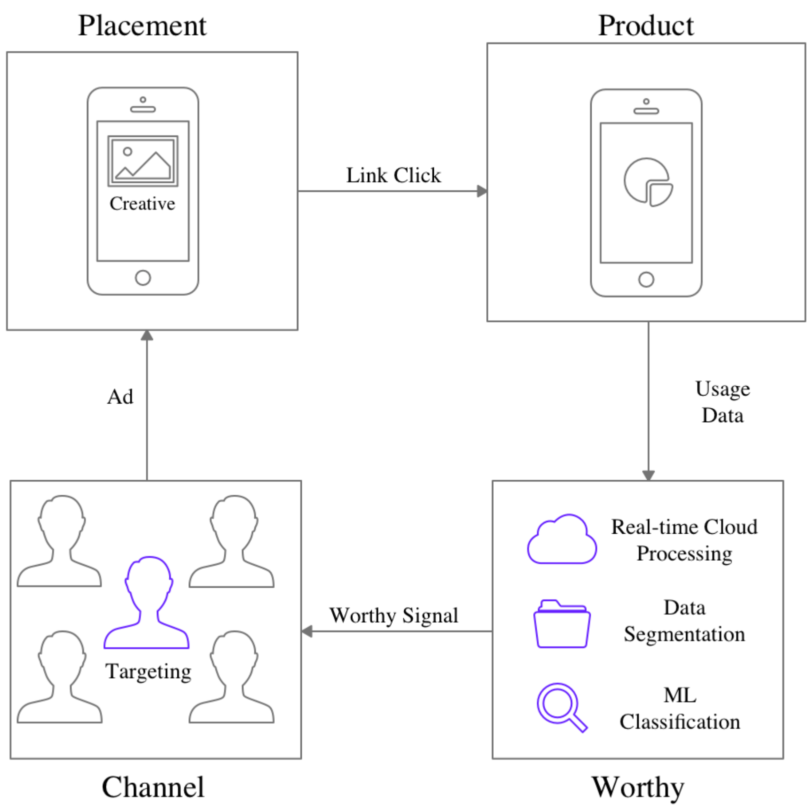


Figure 1: The Advertising Cycle With Worthy

5.1 Worthy Signals Give You An Advantage

To be successful at signal testing you need to be successful at creating powerful signals. Without Worthy, an advertiser would need to hire an internal engineering team to create the cloud infrastructure needed to process and analyze vast amounts of real-time data.

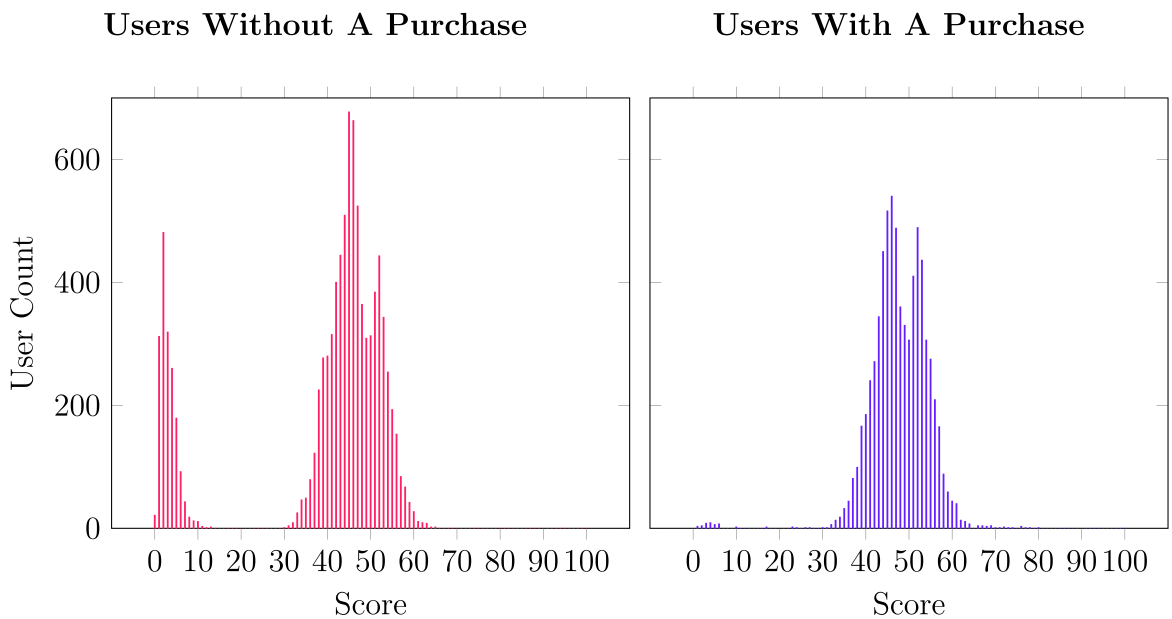
With Worthy, an advertiser doesn't need to focus on building an internal engineer team for signal testing. Worthy's engineers work closely with each advertising partner to create, iterate, and deploy powerful signals.

Developing powerful signals starts with the user journey. Worthy's engineers learn about the unique user journey specific to each advertising partner, understanding key design decisions and each different type of user event. Then, Worthy engineers craft a unique set of features for the specific product incorporating well-known events, like "app opened", with custom built events like "time between install and 10 events".

A set of features enables Worthy engineers to train a machine learning model predicting an advertisers North Star. For example, a North Star may be a subscription or purchase event. Training a machine learning model is unique to each advertising partner. Worthy's engineers evaluate multiple machine learning approaches and fine tune model hyperparameters to find a model that will produce a high performance signal.

After training, Worthy's engineers evaluate the model and the model's ability to produce a high quality signal. In this evaluation sample, the graphs represent distributions of users where the red graph shows users that did not purchase and the green graph shows users that did purchase. The x-axis is a score, 0 to 1, that the machine learning model assigns to a user where 1 indicates a high propensity to purchase.

This model evaluation shows significant signal potential, as there are many red users that can be identified with the following rule: any user with a predicted score below 0.2 is almost certainly red. Since red users do not purchase, this machine learning model could be used to create a signal to optimize for higher quality users with a larger propensity to purchase.



After refining a machine learning model, Worthy engineers deploy a signal into Worthy's signal testing platform. Worthy's cloud infrastructure monitors each advertising partner's data stream in real-time looking for key user events, and fires the signal for all users that meet the signal's criteria.

Advertising partners can see each signal in the Worthy dashboard, along with real-time monitoring charts:

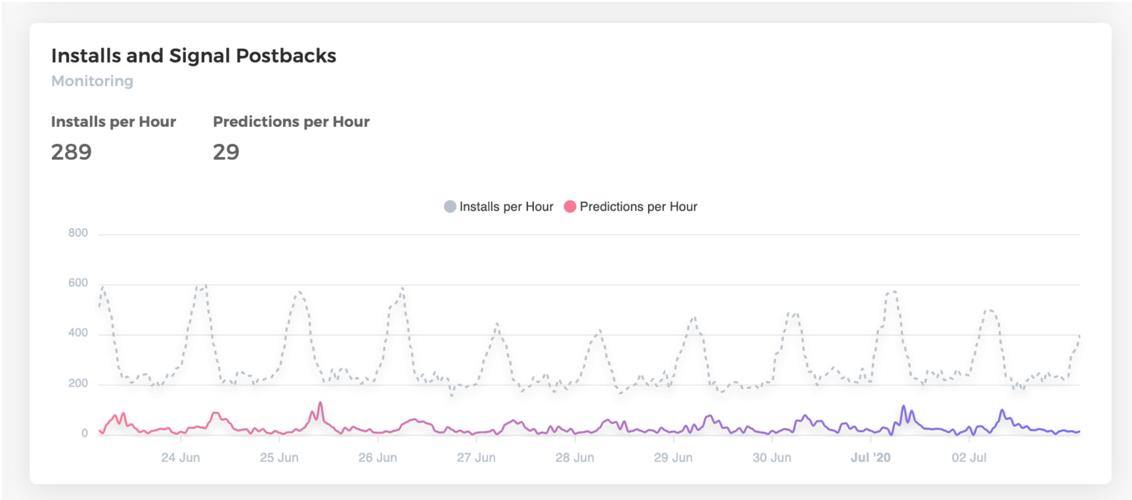


Figure 3: Worthy Dashboard, Signal Monitoring

Signals created by Worthy engineers offer distinct advantages, improving an advertiser's ability to achieve the best ad optimization:

1. **Customized.** Worthy's machine learning models are custom-trained and fine-tuned for each advertising partner by machine learning engineers on Worthy's team. Each signal is specific to the app and advertising partner, giving the signal superior performance.

2. **Dynamic.** Worthy's cloud infrastructure enables Worthy's team to deploy custom signals for each advertising partner at any desired cadence. A signal may fire once a fixed interval of time after install, or a signal may fire multiple times on a daily basis. For example, a signal could fire an event 4 hours after install for the top 50% of users likely to purchase; and then fire another event at the 24 hour mark for the top 10% of users likely to purchase; and then fire another event at the 48 hour mark for the top 5% of users likely to purchase.

3. **Early.** The Worthy event can fire closer to the time of ad click relative to standard user events. While it might take a week for a new user to mature into a paying user, Worthy can predict the behavior within the first day. Instead of waiting a week for the optimization event to appear,

the algorithm now needs only wait minutes. This allows the algorithm to get a quicker feedback signal, which feeds a faster learning cycle.

4. **Value-Adjusted.** The Worthy event is able to adjust for differences in lifetime value. Instead of paying a flat cost per paying user, you can use Worthy to train Facebook to calibrate your bid based on predicted lifetime value.

6 Worthy Data

6.1 Parx

The winning Worthy signal (Worthy Whale Depositor) drove 34% lower cost per First Deposit than the previous best performing signal (First Deposit).

Parx sought to continue acquiring high-value installs in Pennsylvania, but wanted a technique for significantly reducing top-of-funnel acquisition costs without sacrificing install quality and maintaining a high install-to-first-deposit rate.

Worthy helped Parx achieve this by developing a predictive model for deposit behavior. Once the model was developed, Worthy created two different optimization signals for Parx to test.

1. **Worthy (Whale).** Fires after 5 hours for users who are likely to deposit at least $200 within 8 days of install.

2. **Worthy (Standard).** Fires after 5 hours of install for users that are likely to make at least 1 deposit within 8 days of install.

Both Worthy-optimized campaigns drove substantially lower cost per first deposit than the campaign optimizing for the lowest cost per first deposit!

The Worthy (Whale) event was the clear winner due to an extreme CPM discount. It drove 34% lower cost per first time deposit relative to first depositor optimization.

Optimization Event	Cost per 1000 Impressions	Cost per Link Click	Cost per App Install	Cost per New Subscription
Worthy (Whale)	**0.75x**	**0.81x**	**0.77x**	**0.85x**
Worthy (Standard)	0.99x	0.95x	1.13x	0.96x
First Deposit	1.73x	1.53x	1.23x	1.29x
Index	**1.00x**	**1.00x**	**1.00x**	**1.00x**

Table 1: Parx User Acquisition With Worthy

6.2 A Top-Grossing Meditation App

The winning Worthy signal (Worthy High Threshold) drove 26% lower cost per subscription than the previous best performing signal (New Subscription).

A top-grossing meditation app sought to develop a signal that would coax the algorithm toward targeting high quality users without paying a large CPM premium, while optimizing for latent subscriptions which occur outside of the standard optimization window.

Worthy helped a top-grossing mediation app achieve this by predicting a user's propensity to subscribe within 30 days and then:

(a) firing the initial signal earlier in the user lifecycle than the normal subscription event and

(b) passing a stream of signals over time as users convert.

Worthy developed two unique signals for a top-grossing meditation app: a medium and high threshold subscription prediction event.

1. **Worthy Medium Threshold.** Fires 4/24/48 hours after install for users with a relatively high likelihood to subscribe within 30 days, based on re-running the predictive model at each time interval.

2. **Worthy High Threshold.** Fires 4/24/48 hours after install for users with a very high likelihood to subscribe within 30 days, based on re-

running the predictive model at each time interval.

Worthy signals did not drive the lowest CPM, CPC or CPI (install optimization did), nor did they drive the highest subscribing installs (subscription optimization did).

Despite this, Worthy drove the most efficient performance because Worthy was able to acquire high value installs at a significant discount relative to direct subscription optimization. The High Threshold signal drove a 26% lower cost per subscription than direct subscription optimization.

Optimization Event	Cost per 1000 Impressions	Cost per Link Click	Cost per App Install	Cost per New Subscription
Worthy High Threshold	1.12x	1.12x	1.00x	**0.71x**
Worthy Medium Threshold	0.96x	1.00x	0.97x	0.89x
Subscription	1.86x	1.92x	1.77x	0.96x
Content View	0.82x	0.86x	0.98x	1.21x
Install	**0.77x**	**0.71x**	**0.72x**	1.68x
Index	**1.00x**	**1.00x**	**1.00x**	**1.00x**

Table 2: User Acquisition With Worthy For A Top-Grossing Meditation App

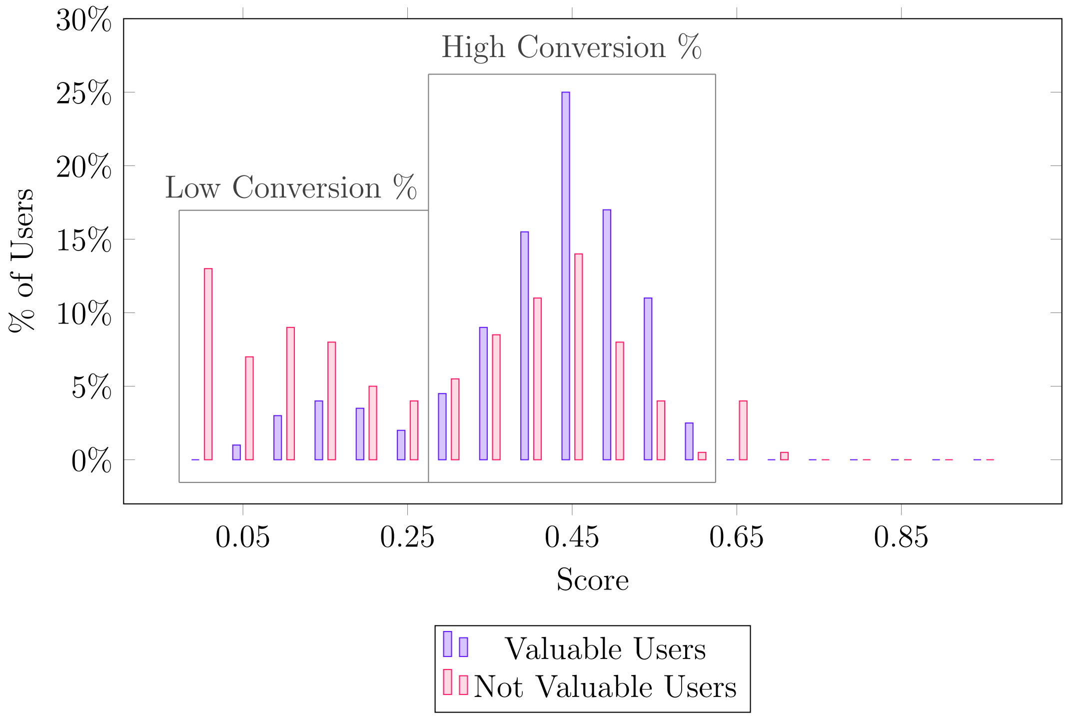
6.3 Abide

The winning Worthy signal drove 17% lower cost per subscription than the previous best performing signal (free trial).

Abide sought to develop a signal that would allow it to optimize for free trials that were likely to subscribe on day 7, rather than the lowest cost per free subscription.

Worthy helped Abide achieve this by predicting the propensity of a trial to subscribe after the 7-day free trial window. Within 2 hours of starting a trial, Worthy was able to flag the majority of trials that would churn before subscribing. Worthy would only fire a signal to Facebook when a new trial generated a score above .40 (in the 'High Likelihood to Subscribe' box below).

Abide Worthy Signal Score Distribution



In the split test, the Worthy event drove an insignificant decrease in cost per trial vs. free trial optimization. It won by improving the conversion quality of trials, reducing cost per subscription by 17%.

Optimization Event	Cost per 1000 Impressions	Cost per Link Click	Cost per App Install	Cost per Trial	Cost per New Subscription
Worthy Event	**0.87x**	**0.93x**	**0.93x**	**0.98x**	**0.83x**
Trial Event	1.18x	1.08x	1.08x	1.02x	1.10x
Index	**1.00x**	**1.00x**	**1.00x**	**1.00x**	**1.00x**

Table 3: Abide User Acquisition With Worthy