



Driving Subscriptions Through User Behavior Modeling and Prediction at Bloomberg Media

Rishab Gupta^(✉), Rohit Parimi, Zachary Weed, Pranav Kundra, Pramod Koneru, and Pratap Koritala

Bloomberg LP, New York City, NY 10022, USA
{rgupta296, rparimi}@bloomberg.net

Abstract. In May 2018, Bloomberg Media launched a digital subscriptions business with a paywall. This meant that users would now be blocked after reading a fixed number of articles per month on Bloomberg.com and the Bloomberg mobile application. For the launch, all users were allowed to read 10 articles for free per month and were offered the same introductory price for subscription. Although we were able to get a large number of subscribers immediately after launch with this “one size fits-all” approach, the conversion rate steadily dropped in the next few months (the initial spike in subscriptions can be attributed to the purchase of subscription by our loyal users). In order to overcome this problem and to keep our subscription conversion numbers high, our team leveraged user engagement data and built a subscription propensity machine learning model to intelligently target users with the right introductory offers. Furthermore, our team used the engagement data trends across months along with the subscription propensity scores to optimize the paywall height for each user, thus maximizing revenue.

In this talk, we will present details about the challenges we faced and approaches we considered to implement the subscription propensity model. We will also talk about how we combined this model with user engagement trends to dynamically calculate the paywall height for each user for every month. Finally, we will share details on how we integrated and used the subscription propensity and dynamic paywall techniques in production, along with results that demonstrate the success of these approaches in increasing Bloomberg Media’s subscription revenue.

Keywords: Machine learning · Data science · User behavior modeling and prediction · Subscription propensity · Elastic paywall · A/B testing

1 Introduction

At the time of the launch of Bloomberg Media’s subscription business, all users of Bloomberg’s web and mobile applications were given the same introductory price and paywall height. Paywall height is the number of articles a user is allowed to read for free per month. However, this is not an optimal approach as content consumption patterns and price sensitivities are different across users and geographical regions. For example, some users are avid news readers and can be considered our core audience whereas some others

are casual readers. Therefore, it is important for us to identify and categorize our users so that we can market the right introductory price to them and maximize subscription revenue. Similarly, having a static paywall height for all users is also not optimal; if the paywall height is too large, we will only be able to block a small percentage of our audience, thereby not providing an incentive to subscribe leading to a loss in subscription revenue and if the paywall height is too small, most users may be prone to leave our site and not come back as it affects their news reading experience. This leads to a loss in Ad revenue as we lose a lot of page views by blocking content and from users who never return. Therefore, the right thing to do is nurture a user's engagement so that we can target them at an opportune moment while ensuring the Ad revenue is not negatively affected.

In order to target users with appropriate introductory offers, we decided to build a subscription propensity machine learning model which classifies our users into two buckets: users with high propensity to subscribe and users with low propensity to subscribe. The input to the model is user clickstream data which contains information about user's interactions with our content along with other metadata. The output is the propensity score (or the probability to subscribe) for that user. More details about this problem and our solution are explained in Sect. 2.

Once we had subscription propensities for users, our goal is to optimize the paywall height for each user such that we maximize subscription and Ad revenue combined. To achieve this, we make use of the subscription propensity scores along with user engagement trends in previous months to compute a paywall height for the current month for each user. More details about the problem and its solution are provided in Sect. 3.

These two optimizations led to a statistically significant increase in conversion rates and led to a total increase in revenue. The subscription propensity model is at the core of the solution. We will talk about some of the A/B test results and our insights in Sect. 4.

Some of the most important yet frequently ignored details when using machine learning models in the real-world are details about data pipelines (data gathering, cleaning and filtering, GDPR handling, and feature generation) and continuous model evaluation and validation in the production environment. Although these details are not part of this paper, we plan to present these details as part of our presentation.

2 Subscription Propensity Model

The input data for our subscription propensity model is user click stream data. This data can provide a lot of information about a user's behavior on our platforms. We can look at their activity with respect to article interactions, which can be broken down by page types. We also get information about the country of access along with device information such as browser and operating system. More details about this will be shared during our talk.

The class label for this problem was constructed by using active subscribers as positive examples and all other users as negative examples. This gave rise to the class imbalance problem, one of the most common challenges when working with real-world problems. For example, in our dataset, the subscribers constituted only about 0.025% of

all users. An important consideration when working with imbalanced datasets is selecting the right evaluation metric. Model accuracy is not an insightful metric to use because the model will have 99.975% accuracy if it predicted every user as a non-subscriber. Therefore, to evaluate our model, we used area under precision-recall curve (AUPRC) along with precision and recall values for the positive and the negative classes.

We experimented with multiple models like logistic regression, neural nets, and decision trees and used k-fold cross-validation to tune hyper-parameters for each of these models. For our problem, Logistic Regression performed the best and is currently used in production. More details on feature engineering, experimentation, and periodic model evaluation in production will be shared during our presentation.

The goal of a classifier is to predict whether a data point belongs to a positive class or a negative class. However, instead of using the predictions from the classifier directly in production, we decided to use the probability scores given by the classifier as the subscription propensity score. These probability scores represent a user's likelihood to subscribe. This allows us to selectively target, for example, high propensity users who are not yet subscribers.

We used the results of the subscription propensity model, more precisely the propensity scores, to run A/B tests on Bloomberg's website. In these A/B tests, we experimented with different introductory offers on targets of users bucketed according to their propensity scores. Each A/B test was run for roughly a month and at the end of the month, the variation that resulted in statistically significant increase in conversions as compared to control was declared the winner and was made the new control for subsequent A/B test. Details about our A/B test set-up and results for the same will be shared in our presentation.

3 Elastic Paywall Optimization

The elastic paywall problem was aimed to determine the optimum paywall height for each user. This problem is central to business revenue since it controls the trade-off between subscriptions and ad revenue. As mentioned earlier, blocking users with a paywall can lead to higher subscription conversions with a loss in ad revenue and vice-versa.

Subscription propensity is a valuable signal since a high value indicates that a user is likely to subscribe and they should be targeted accordingly. This can be treated as a representative for subscription revenue. However, it cannot be used in isolation to decide the opportune moment to affect said targeting. As mentioned earlier, a naive approach of aggressively blocking high propensity users leads to higher bounce rates which has repercussions on both ads and subscription revenue. While we want to increase subscription conversion rates, we need to be mindful that we do not do so at the risk of impacting ad revenue. This calls for the need of additional signals to be used in combination with subscription propensity. A key signal that we identified is user engagement trends, which can serve as a representative for ads revenue, since more on-site engagement leads to higher ad revenue.

The subscription propensity and engagement trends thus provide us with vital insight into both aspects of the ads versus subscription revenue trade-off. In order for a paywall to be truly elastic, it is necessary to consider the prior paywall height instead of starting fresh

each time. The previous paywall height for users can be factored into the calculations to achieve this. The diagram below highlights the factors that go into the elastic paywall height calculation engine. This paywall height is then applied to users for the coming month. As a part of our talk, we will talk about how we solved the elastic paywall optimization problem using these key metrics, our method of testing and some of the observed results (Fig 1).

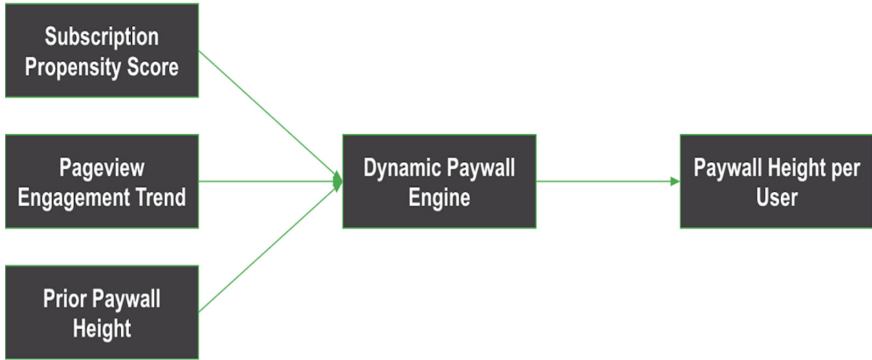


Fig. 1. Block diagram to illustrate inputs that contribute to the calculation of a user’s paywall height for the month

4 Results

We use an in-house A/B testing solution to run experiments on our website. This allows us to create custom target audience segments, set up experiments for those targets and measure experiment performance. Some A/B test results for the subscription propensity model are as follows (Table 1):

Table 1. A/B test results for the subscription propensity model

User segment	Conversion lift	Revenue lift	Statistical significance
High propensity users with 2+ paywall hits	57%	27%	98.3%
Low propensity users - Segment 1	125%	186%	98.8%
Low propensity users - Segment 2	144%	150%	97.8%

Similarly, we ran A/B tests to experiment with the elastic paywall. We tested a few variants of the elastic paywall with respect to how the paywall adjusted itself depending

Table 2. A/B test results for Elastic Paywall

Segment	Conversion Lift (vs. height of 10)	Revenue Lift	Change in ad revenue	Statistical significance	Change in paywall stop rate
High Propensity users with 0 paywall hits	+108%	+20%	+4%	99.6%	+114%
High Propensity users with 1+ paywall hits	+114%	+25%	+2%	99.9%	+36%
Low Propensity users with 0 paywall hits	+33%	+12%	+0%	96.4%	+140%
Low Propensity users with 1+ paywall hits	+108%	+22%	-12%	95.2%	+21%

on subscription propensity and article engagement. Following are some key results we observed (Table 2):

The propensity experiment numbers clearly demonstrate the success of using subscription propensity to target users effectively with different introductory offers. The elastic paywall numbers show that some variants were more successful than others in increasing conversions and overall revenue. One key thing to note is that there wasn't a major impact on ad revenue by this effort while there was a lift in conversions and the paywall stop rate increased across the board. These efforts allowed Bloomberg Media to exceed its year-end subscriptions goal by about 40% and were a big success story.