The Workshops of the Thirty-Second
AAAI Conference on Artificial Intelligence

# Bingeability and Ad Tolerance: New Metrics for the Streaming Media Age

**Prashant Rajaram**
prajaram@umich.edu
University of Michigan

**Puneet Manchanda**
pmanchan@umich.edu
University of Michigan

**Eric Schwartz**
ericmsch@umich.edu
University of Michigan

## Abstract

Binge-watching TV shows on streaming services is becoming increasingly popular. However, there is a paucity of comprehensive metrics to effectively summarize such media watching behavior. We address this gap by presenting two new metrics—Bingeability and Ad Tolerance—to quantify key aspects of watching streaming TV interspersed with ads. These metrics are motivated by consumer psychology literature on hedonic adaptation and also reflect media consumption behavior. Using machine learning methods, including ensembles of classification trees, we identify the key predictors of these metrics, study non-linear effects, and rank the predictors in order of predictive power. The superiority and validity of these metrics is also discussed.

## 1. Introduction

Binge-watching TV shows on streaming services is becoming increasingly popular (Holloway 2016). Binge-watching refers to rapidly viewing multiple episodes of the same TV show (series) such that the user can self-schedule the amount of time spent watching content (Oxford Dictionary 2017, Jenner 2015). Popular streaming services have more qualified interpretations of binge-watching. TiVo defines it as viewing more than 3 episodes of a TV show in one day (TiVo 2015) whereas Netflix conducted a poll and found that its users perceive watching 2 to 6 episodes of a TV show in one sitting as binge-watching (West 2013). On the other hand, Ameri, Honka, and Xie (2017) consider watching more than 3 hours on average per day to finish a season of a show as binge-watching. As can be seen from the above, there is little consensus on what constitutes binge-watching, especially with respect to the duration of a sitting, the number of episodes to be seen in a sitting and whether time spent between two sittings should be considered. Moreover, there is no talk about how long an episode should be, whether episodes can be watched partially or out of sequence, and whether content can be fast-forwarded. This suggests that the definition of binge-watching is still evolving.

Past work in the marketing literature has looked at how an increase in ad exposure discourages binge-watching behavior in the same sitting (Schweidel and Moe 2016). From a marketer's perspective, an open question is whether past viewing activity on the platform can be used to predict future viewing activity and user tolerance for ads.

The goal of our paper is to (a) develop two new metrics that quantify key aspects of watching streaming TV interspersed with ads, (b) explore the interplay between the metrics, (c) show the superiority and validity of these metrics and explain their relevance for any media platform that streams content along with ads, and (d) delineate the predictors of these metrics via the use of Machine Learning methods. Our first metric is "Bingeability", or the count of the 'complete unique episodes' watched in a session. It represents the count of episodes that contribute towards binge-watching behavior. Our second metric is "Ad Tolerance" that is defined as a measure of the willingness of a user to continue watching content after seeing ads in a session.

## 2. Data

Our data are from the streaming provider Hulu for a random sample of 1000 users for the period Feb 28, 2009 to June 29, 2009. At this time, Hulu only offered a free streaming of content interrupted by ads. While Hulu has moved on to paid options, this model has remained popular with other extant streaming services such as TubiTV, Crackle and Popcornflix. As binge-watching exists primarily for TV shows, we focus our analysis on titles that are TV shows. A 'session' (or sitting) is defined as time spent watching show content or ads from exactly one TV show separated by 60 minutes or more of inactivity. The time separation of 60 minutes is

---

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

consistent with the usage in Schweidel and Moe (2016). A session can be split into the following parts:

$$\text{Session Time} = \text{Content Time} + \text{Ad Time} + \text{Filler Content Time} + \text{Pauses} - \text{Fast Forward} + \text{Rewind} \qquad (1)$$

where, Session Time represents Calendar Time. Content Time is time spent viewing show content (including content skipped in fast-forwards but excluding content seen again in rewinds), Ad Time is time spent viewing ads, and Filler Content Time is time spent watching content such as interviews with the star cast. The exact value of each of these variables is available in our panel data. In addition, there are unmeasured variables that complete the above equation—Pauses is the time spent in a break, Fast Forward is the duration of content fast-forwarded, and Rewind is the duration of content rewound. We only select those users whose frequency of visits to the platform to watch TV shows span a calendar period of more than one week. This is done to ensure that the shortlisted user base has a minimum level of engagement with TV shows on the Hulu platform. As a result, we are left with a pool of 476 users who watch 388 shows across 12,309 sessions.

## 3. Metrics

### 3.1 Bingeability

Bingeability for a session represents the effective count of episodes that contribute towards binge-watching behavior. We define it as a count of the complete unique episodes watched in a session. A unique episode is counted only if the following conditions are met:

$$Bingeability = \sum_{i=1}^{n} \mathbb{1}\left\{\begin{matrix} Content\ Length_i - 5\ mins \le\ Content\ Time_i \le \\ Calendar\ Time_i - Ad\ Time_i \end{matrix}\right\}$$

where, $\mathbb{1}$ is an indicator function, $i$ denotes a unique episode, $n$ is the number of unique episodes watched in a session, $Content\ Time_i$ is the time spent watching content for episode $i$, $Content\ Length_i$ is the length of episode $i$ including opening and end credits, 5 mins is an upper bound on the combined duration of opening and end credits in an episode, $Calendar\ Time_i$ is the clock time spent and $Ad\ Time_i$ is the time spent watching ads (ads cannot be fast-forwarded, rewound or skipped). We explain the two conditions in the indicator function below.

**Skipping:** $Content\ Time_i \ge\ Content\ Length_i - 5\ min$
The sum of opening and end credits for TV shows are generally less than 5 minutes (ABC 2014, Ingram 2016) which can be considered a conservative upper bound. This is subtracted from $Content\ Length_i$ as users are less likely to watch credits when they are binge-watching the show (Nededog 2017, Miller 2017). After subtracting the maximum possible time involved in opening and end credits (5 mins) from $Content\ Length_i$, if the difference remains greater than $Content\ Time_i$, then we can conclude that the user has left the episode unfinished. The act of 'skipping' (not watching) content (excluding credits) in the end portion of an episode is not considered as binge-watching in our analysis. We explain this below.

Monotonically increasing patterns of utility have been discussed in the literature on sensitization where utility or arousal levels are proposed to increase with time from repeated exposure to a stimulus with positive valence (Frederick and Loewenstein 1999, Nelson et. al. 2009). If a user's utility level begins to decrease from watching show content, then the user is unlikely to continue watching. Hence, we assume that binge-watching could be primarily a consequence of experiencing a monotonic increase in utility with each passing moment of the viewing experience. Skipping, on the other hand, maybe indicative of dissatisfaction with the content being viewed and a decision to avoid a decrease in utility from continued watching of show content.

**Fast-forwarding:** $Content\ Time_i \le Calendar\ Time_i - Ad\ Time_i$
There may be occasions when the user chooses to excessively fast-forward certain portions of an episode. This would result in a greater increase in $Content\ Time_i$ than $Calendar\ Time_i$. Fast-forwarding, like skipping, is generally indicative of dissatisfaction with the available content and is a decision to avoid a further decrease in utility from continued watching. Hence, Condition 2 seeks to avoid considering episodes in which a user engages in excessive fast-forwards to count towards the Bingeability metric. Another way to think about fast-forwarding behavior is that it is a consequence of the inability of the content to engage the user. This signals a lack of binge-worthiness of the content.

### 3.2 Ad Tolerance

We define Ad Tolerance of a session as a measure of the willingness of a user to continue watching content after encountering ads in a session. It sums the time spent watching an ad and the content following an ad, while compensating for the time elapsed since the last ad. The metric is expressed as follows:

$$Ad\ Tolerance = \sum_{j=1}^{n_e}\sum_{i=1}^{n_{aj}}(AdDuration_{ij} + ConEnd_{ij} - CalAd_{ij})$$

where, $j$ is an episode, $n_e$ is number of episodes watched in a session, $i$ is an ad group (sequence of consecutive ads), and $n_{aj}$ is number of ad groups watched in episode $j$. $AdDuration_{ij}$ is the duration of ad group $i$ that is watched in episode $j$, $ConEnd_{ij}$ is content watched till the end of the session after watching ad group $i$ in episode $j$, $CalAd_{ij}$ is calendar time from the end of the previous ad group in the same session till the beginning of ad group $i$ in episode $j$. There are two main assumptions made in the construction of this metric: (1) Apart from ads, there are no exogenous factors that influence the user to end a user-session (2) Ads of the same duration are interchangeable.

When a user sits through an ad group, the ad tolerance of the user increases which is captured by $AdDuration_{ij}$, the duration of the $i^{th}$ ad group that is watched in the $j^{th}$ episode. While a user does not have the option to fast-forward, rewind or skip ads, a user can partially watch an ad by exiting the session in the middle of the ad or skipping to the next episode in sequence. The second term on the right-hand side of the metric, $ConEnd_{ij}$, measures the time spent watching content in the session after watching ad group $i$ in episode $j$. Longer durations suggest higher tolerance for the previously seen ad.

The third term on the right-hand side of the metric, $CalAd_{ij}$ measures the calendar time from the end of the previous ad group in the same session till the beginning of ad group $i$ in episode $j$. It is subtracted from the sum of the previous two terms because it accounts for the time available for sensitization or adaptation to show content and the absence of ads (Frederick and Loewenstein 1999, Kahneman 2003, Nelson, Meyvis, and Galak 2009). While each successive value of $ConEnd_{ij}$ double counts part of the content time that is remaining till the end of the session, $CalAd_{ij}$ compensates for it.

### 3.3 Summary Statistics

There are 1,237 sessions in which ad exposure was not registered in the data. We do not directly analyze such sessions because we cannot measure Ad Tolerance for them. Table 1 shows the summary statistics of the two metrics. Though the unit of Ad Tolerance is minutes, it cannot be directly interpreted as time spent on some activity. It is a scale of the 'willingness to watch content after ads', and hence can take negative values.

It is important to show that the two proposed metrics are related but distinct. From Figure 1, it can be seen that the relationship is quadratic in nature. The Pearson correlation coefficient between the metrics is 0.81 over the full range of the data and is 0.66 over the 95-percentile range. Thus, using both metrics is likely to be useful.

### 3.4 Superiority, Validity and Relevance

The proposed Bingeability metric is more comprehensive and complete relative to analogous measures proposed, especially in practice. For example, one of the most commonly accepted metrics is the one proposed by Netflix that measures the count of unique episodes watched of the same TV show in one session. However, this metric does not consider whether users are watching each episode completely and/or displaying excessive fast-forwarding behavior. Presence of either of these behaviors indicates dissatisfaction with the content (episode). We carefully establish boundary conditions in our metric that take cognizance of these behaviors. Note that the proposed metric and the Netflix metric are correlated, with a correlation coefficient of 0.91 over the full range and 0.70 over the 95-percentile range of our metric. This provides face validity for our metric while also allowing for it to capture novel information (Ailawadi, Lehman, and Neslin 2003). For the second proposed metric, Ad Tolerance, there are no comparable metrics in practice or academic research to the best of our knowledge.

| | Bingeability (count) | Ad Tolerance (minutes) |
|---|---|---|
| Min | 0 | -117.68 |
| 2.5% | 0 | -2.29 |
| Median | 1 | 48.66 |
| 97.5% | 5 | 638.96 |
| Max | 57 | 35,749.31 |

*Table 1: Metric Summary Statistics*

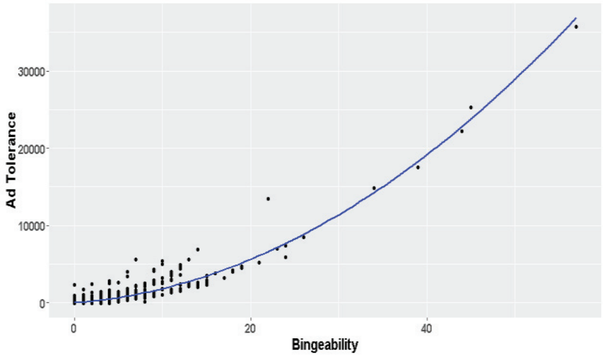

*Figure 1: Ad Tolerance vs Bingeability*

Both metrics are likely to be highly relevant to streaming providers. The high level of granularity of these metrics (at the user-session level) could help streaming providers develop recommendation mechanisms promoting the right show at the right time to the right user. All else being equal, providers should recommend shows with a higher expected product of Bingeability and average episode length, so that users actually watch more of that show leading to higher engagement with the platform. In addition, providers can also predict which sessions will have high Ad Tolerance, allowing them to show the right number of ads in such sessions without decreasing engagement levels, thereby maximizing advertising revenue.

## 4. Feature Generation

Given that these metrics are likely to be useful and relevant, we now turn to describing the predictors of Bingeability and Ad Tolerance. The main question that we ask is as follows: given a user's decision to begin watching a specific TV show at a specific time, can we predict her Bingeability and Ad Tolerance levels from her historical one-week activity on Hulu. We divide the feature generation process into 3 steps:

### A. Current Variables:

Values of these variables are known in the present. There are 892 current variables which include fixed effects for user, show, genre, month, week, day and time of day.

### B. Functions for Watching only TV Shows:

We adopt an approach similar to Yoganarasimhan (2016) to generate features by constructing functions. These functions are implemented over the one-week historical viewing activity for TV shows for the said user before the current session commences. We generate functions that vary with day and time of day to explore whether experiences that occur at a particular time in the past are significant predictors of Bingeability and Ad Tolerance. There are 12 functions that generate 92 features. For example, *Bingeability Sum (Show, Day, Time of Day (TOD))* calculates the sum of Bingeability of a user over the past 1 week for the **Show** she is about to watch over that **Day** at that **Time of Day.** We consider a **Day** as a Weekend or a Weekday and a **Time of Day** as one of the five: Early Morning: 7–10am, Day Time: 10am–5pm, Early Fringe: 5pm–8pm; Prime Time: 8pm–11pm, Late Fringe: 11pm–7am (Schweidel and Moe 2016). A total of 8 features can be generated by the same function when any of the variables in brackets are replaced with *any Show, any Day* or *any Time of Day*.

### C. Functions for Watching TV Shows or Movies:

These functions also consider historical one-week sessions in which movies were seen. This is important as past advertising stimuli while watching movies on the platform could influence the decision to see a TV show in the current session. There are 7 functions that generate a total of 120 features. For example:

*Ad Count (Ad Length, Title, Day, TOD)* calculates the number of ads of length **Ad Length** shown to the user for that **Title** over that **Day** at that **Time of Day**. Based on the distribution of Ad Length, we divide it into 4 categories: 1 (1–13sec), 2 (13–26 sec), 3 (26–39 sec) and 4 (>39 sec).

*Ad Diversity (Title, Day, TOD)* finds the average percentage of diverse ads per session shown to the user for that **Title** over that **Day** at that **Time of Day.** We use a combination of Ad Industry (16 categories such as CPG, Telecom, etc.) and Ad Length (4 categories) to generate 64 ad combinations. As we do not have unique identifying information for each ad, we cannot measure whether an ad is truly unique.

## 5. Model

The purpose of our model is twofold: To identify the predictors and to infer the direction of their average effect on Bingeability and Ad Tolerance. We propose the Elastic Net, an extension of the LASSO, to meet our goals. Tibshirani (1996) introduced the LASSO which is a penalized regression with the constraint expressed as an $L_1$ norm of the coefficients:

$$\hat{\beta} = arg\,\underset{\beta}{min}\,\frac{1}{2}(y - X\beta)'(y - X\beta) + \lambda\|\beta\|_1 \qquad (2)$$

where, $y$ is the outcome variable, X is the matrix of normalized regressors, and $\lambda > 0$ is a tuning or penalization parameter that helps minimize variance. However, the LASSO is unable to select an entire group of predictors when there is multicollinearity among them (Kyung et al. 2010). The Elastic Net proposed by Zou and Hastie (2005) solves the problem of selecting an unknown group of variables that are multi-collinear (Kyung et al. 2010). In our situation, there are 19 functions that generate 212 features, many of which have high collinearity. The Elastic Net provides equal weight to the coefficients in a group of multi-collinear predictors and hence either selects or rejects them as a group. It is an extension of the LASSO as it is formed by adding an $L_2$ norm of the coefficients to equation (2). We use the general form of the Elastic Net proposed by Hastie, Tibshirani, and Wainwright (2015):

$$\hat{\beta} = arg\,\underset{\beta}{min}\,\frac{1}{2}(y - X\beta)'(y - X\beta) + \lambda\left[\alpha\|\beta\|_1 + \frac{1}{2}(1-\alpha)\|\beta\|_2^2\right] \qquad (3)$$

where, $\lambda$ is the tuning parameter and $\alpha$ is kept at 0.5. When $\alpha = 0$, equation (3) will reduce to the form for Ridge Regression. When $\alpha = 1$, equation (3) will reduce to the form for LASSO.

We run two models, one where the outcome is Bingeability which is a count variable, and the other where the outcome is Ad Tolerance which is a continuous variable. In the first case, we assume the outcome to have a Poisson distribution, and hence apply a log-link on the outcome variable. In the second case, we assume the outcome to be normally distributed. A limitation of the Elastic Net is that even for fixed values of the tuning parameter, it is difficult to estimate the standard errors of the estimate $\hat{\beta}$. One approach to estimate the standard errors is to run a Poisson regression and linear regression on the variables with non-zero coefficients selected by the Elastic Net when the outcome is Bingeability and Ad Tolerance, respectively (Friedman, Hastie, and Tibshirani 2001). The parameters identified from this approach will have less bias but more variance. However, we can now find the significance of the parameter values and compare the sign of the significant parameters identified from the regression with those from the Elastic Net. Finding consistency in the sign presents a much stronger argument for the direction of the predicted effect.

## 6. Results

We randomly divide the set of 476 users into a training sample (80%) and test sample (20%). As mentioned in Section 3.3, we remove the sessions with no registered ad exposure from the training and test sample. We are left with a training sample of 8,655 sessions and a test sample of 2,417 sessions. Our method is a strict procedure that trains information obtained from one group of people to predict performance of a different group. The model may also be used to estimate the metrics for a new show. The tuning parameter of the Elastic Net is selected by cross-validation. The MSE for the estimates obtained from the Elastic Net model are compared with other models in Table 2. When the outcome is Bingeability, the MSE is the lowest for Ridge Regression which indicates that the covariates are good linear predictors of the outcome variable. When the outcome is Ad Tolerance, the lowest MSE is for XGBoost (discussed in Section 7) and the highest MSE is for Ridge Regression. This indicates that the covariates are not good linear predictors of Ad Tolerance.

| | Bingeability | Ad Tolerance |
|---|---|---|
| Elastic Net | 1.275 | 64,959 |
| LASSO | 1.276 | 63,646 |
| Ridge Reg | 1.268 | 76,186 |
| XGBoost | 1.279 | 63,459 |

*Table 2: Model Comparison*

### 6.1 Feature Selection

The Elastic Net selects 286 predictors when the outcome is Bingeability and 9 predictors when the outcome is Ad Tolerance. We now run a Poisson/linear regression on only the variables with non-zero coefficients selected by the Elastic Net. When the outcome is Bingeability there are 129 parameters that are significant at the 5% level. When the outcome is Ad Tolerance, all the 9 parameters are found to be significant at the 5% level. All the 138 parameters have a matching sign across both Elastic Net and Poisson/linear regression. We display the direction of influence of the predictors (except user, show and genre fixed effects) on the outcome using a + or – sign in Tables 3a and 3b. The expansion for each abbreviation used is mentioned in Table 4.

**6.1.1 Bingeability**

Among the past predictors, higher sum of Ad Tolerance (ATS) while watching the same TV show in the past week is predictive of lower Bingeability in the present session. Having completely watched at least one episode of the same TV show on the same Day in the past week (BI) is predictive of higher Bingeability in the present session.

**6.1.2 Ad Tolerance**

ATS (same Title, same Day, same TOD) is a predictor of lower Ad Tolerance in the current session. This indicates that, all else being the same, Ad Tolerance levels for a user corresponding to the same Title watched on the same Day at the same Time of Day, decrease on average with time.

| No | Feature Summary | Effect |
|---|---|---|
| 1 | Weekday | – |
| 2 | Time of Day – Early Morning | – |
| 3 | ESC (2 features) | – |
| 4 | AC (4 features) | + |
| 5 | AP (any Title, same Day, same TOD) | – |
| 6 | AS (2 features) | + |
| | AS (same Title, any Day, any TOD) | – |
| 7 | ADSC (1-13s, any Title, any Day, same TOD) | – |
| 8 | ATS (same Title, same Day, same/any TOD) | – |
| | ATS (same Title, any Day, same TOD) | – |
| 9 | BI (same Show, same Day, any TOD) | + |
| 10 | BSC (any Show, same Day, any TOD) | + |

*Table 3a: Influence of the features on Bingeability*

| No | Feature Summary | Effect |
|---|---|---|
| 1 | AC (any AL, same Title, any Day, any TOD) | + |
| 2 | AP (same Title, any Day, same TOD) | + |
| 3 | AS (any Title, same Day, any TOD) | + |
| 4 | ATS (same Title, same Day, same TOD) | – |

*Table 3b: Influence of the features on Ad Tolerance*

| Abbreviation | Expansion |
|---|---|
| AC | Ad Count |
| ADD | Ad Diversity |
| ADSC | Ad Session Count |
| AP | Ad Proportion |
| AS | Ad Stop |
| ATS | Ad Tolerance Sum |
| BI | Bingeability Indicator |
| BS | Bingeability Sum |
| BSC | Bingeability Session Count |
| ESC | Episode Session Count |

*Table 4: Expansion of Abbreviations*

## 7. Tree-based Boosting Methods

We resort to tree-based Boosting methods to study non-linear effects of the predictors and to identify the important predictors. Allowing higher order interactions between the predictors improves prediction accuracy for non-linear models. These interactions are captured when the regression tree makes more than one split on a branch at the same time, with each additional split allowing for an increase in the degree of interaction between the predictors (Breiman 1984). Automatic variable selection is performed, and the selected features can be ranked in order of predictive power. Boosted

Regression Trees are a weighted linear combination of regression trees, with each tree trained greedily in sequence to improve the final output (Friedman, Hastie, and Tibshirani 2000). This output can be presented as follows:

$$F_N(x) = \sum_{i=1}^{N} \alpha_i f_i(x, \beta_i)$$

where, $f_i(x, \beta_i)$ is the function modelled by the $i^{th}$ regression tree, and $\alpha_i$ is the weight associated with it. Both $f_i$ and $\alpha_i$ are learnt over the training sample. $f_i$ is chosen such that it minimizes a loss function—least squares error in linear regression and negative log likelihood in Poisson regression. At each step, gradient descent computes the new value of $f_i$ that minimizes the average value of the loss function:

$$F_i(x) = F_{i-1}(x) - \gamma . g_i$$

where $g_i$, with components $g_{ji} = \left[\frac{dL(y_j, F(x_j))}{dF(x_j)}\right]_{F(x_j)=F_{i-1}(x_j)}$, is the gradient of the Loss function $L(y, F(x))$ evaluated at $F(x) = F_{i-1}(x)$, and $\gamma$ is the step length.

### 7.1 Extreme Gradient Boosting

We adopt a recent extension of Gradient Boosting called Extreme Gradient Boosting (XGBoost) (Chen and Guestrin 2016). It allows for regularization to prevent overfitting. Regularization is implemented by selecting the threshold for loss reduction required to make a further partition on a leaf node of the tree. We choose the optimal value of the threshold and number of boosting rounds by cross validation.

The MSE of the predictions are compared in Table 2. XGBoost has the poorest performance when the outcome is Bingeability. This shows that Bingeability indeed has a strong linear relationship with most of its predictors. When the outcome is Ad Tolerance, XGBoost performs the best, thus showing that the tree based model can capture non-linear effects and help improve prediction accuracy.

### 7.2. Feature Importance

We study the presence of up to three simultaneous splits in the construction of the decision tree to test the importance of higher order interactions. A commonly used metric to measure feature importance is 'Variance Reduction' for Regression Trees. It is the 'Gain' achieved when the tree is split on some covariate. Gain for a covariate is the maximum reduction in Root MSE (in linear regression) or Negative Log Likelihood (in Poisson regression) that can be achieved when the covariate space is split based on some value of the covariate. These are used as evaluation metrics when outcomes are Ad Tolerance and Bingeability respectively. We identify the zero order and higher order covariates that are most frequently split, and report the percentage gain for the said feature out of a maximum of 100 features of that type. The top 3 features are shown in Tables 5a and 5b. The abbreviations used are explained in Table 4 and the numeric code beside each function indicates a distinct feature generated by that function. The features in bold are the covariates of interest that do not include fixed effects.

| Rank | 0 ord | Gain | 1 ord | Gain | 2 ord | Gain |
|---|---|---|---|---|---|---|
| 1 | User1 (fixed effect) | 29.34 | BS2<br>User1 | 7.87 | BS2 BS3<br>User1 | 4.50 |
| 2 | **BS1** | 12.92 | ADD1<br>User1 | 7.25 | BS1<br>Comedy<br>User1 | 3.62 |
| 3 | **BS2** | 5.41 | **ADD1<br>BS1** | 5.90 | BS2<br>User1<br>User2 | 3.48 |

*Table 5a: Top 3 Features for Bingeability*

In Table 5a, the zero order (0 ord) past predictors with the highest gain are features of *Bingeability Sum*. In Table 5b, the most important zero order past predictors are features of *Ad Diversity* and *Bingeability Sum*.

| Rank | 0 ord | Gain | 1 ord | Gain | 2 ord | Gain |
|---|---|---|---|---|---|---|
| 1 | User1 | 30.59 | ADD4<br>User1 | 19.41 | ADD4<br>BS4<br>User1 | 17.54 |
| 2 | **ADD4** | 19.53 | **ADD4<br>BS4** | 15.55 | ADD4<br>ESC6<br>User1 | 17.44 |
| 3 | **BS4** | 17.03 | User3<br>User1 | 14.26 | **ADD4<br>BS1 BS4** | 15.09 |

*Table 5b: Top 3 Features for Ad Tolerance*

## 8. Conclusion

We developed two new metrics—Bingeability and Ad Tolerance—motivated by consumer psychology literature on hedonic adaptation and consumer practices of media consumption. We show that these metrics are novel, comprehensive and likely to be of value to streaming providers. Using machine learning techniques and a historical one-week of viewing activity, we identify the key predictors of these metrics from a set of 1104 predictors. The predictors have mainly a linear effect on Bingeability but a non-linear effect on Ad Tolerance. Using Extreme Gradient Boosting we study non-linear effects, identify higher order interactions among the predictors and rank the predictors in terms of predictive power. In future work, we plan to further analyze the direction and magnitude of the influence of the linear and non-linear predictors on Bingeability and Ad Tolerance.

# References

ABC. 2014. End Credit Guidelines. Retrieved from abc.net.au: http://www.abc.net.au/tv/independent/doc/ABC_Commissioned_Productions_Credit_Guidelines_2014.pdf

Ailawadi, K. L., Lehmann, D. R., & Neslin, S. A. 2003. Revenue premium as an outcome measure of brand equity. *Journal of Marketing*, 67(4), 1-17.

Ameri, M., Honka, E., & Xie, Y. 2016. Word-of-Mouth, Observational Learning, and Product Adoption: Evidence from an Anime Platform. *Working Paper*.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. 1984. *Classification and regression trees*. CRC press.

Chen, T., & Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). ACM.

Frederick, S., and Loewenstein, G. 1999. Hedonic adaptation. in D. Kahneman, E. Diener, and N. Schwartz [eds]. *Scientific Perspectives on Enjoyment, Suffering, and Well-Being*. New York. Russell Sage Foundation.

Friedman, J., Hastie, T., & Tibshirani, R. 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, *28*(2), 337-407.

Friedman, J., Hastie, T., & Tibshirani, R. 2001. *The elements of statistical learning* (Vol. 1, pp. 241-249). New York: Springer series in statistics.

Hastie, T., Tibshirani, R., & Wainwright, M. 2015. Statistical learning with sparsity: the lasso and generalizations. CRC press.

Holloway, D. 2016. *Netflix Looks at Which Shows Are Most Binge-Watched*. Retrieved from www.variety.com: http://variety.com/2016/tv/news/netflix-looks-at-which-shows-are-most-binge-watched-1201791061/

Ingram, K. 2016. *A brief history of TV shows' opening credit sequences*. Retrieved from theweek.com: http://theweek.com/articles/632836/brief-history-tv-shows-opening-credit-sequences

Jenner, M. 2015. Binge-watching: Video-on-demand, quality TV and mainstreaming fandom. *International Journal of Cultural Studies*, *20*(3), 304-320.

Kahneman, D. 2003. Experienced utility and objective happiness: A moment-based approach. *The psychology of economic decisions*, *1*, 187-208.

Kyung, M., Gill, J., Ghosh, M., & Casella, G. 2010. Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*, *5*(2), 369-411.

Miller, L. S. 2017. *Netflix Shouldn't Let Fans Skip Movie Credits, But We'll Allow It For TV Shows*. Retrieved from www.indiewire.com: http://www.indiewire.com/2017/05/netflix-skip-intro-bad-for-film-good-for-tv-1201817946/

Nededog, J. 2017. *Some lucky Netflix members have a cool new 'skip intro' button to make binge-watching better*. Retrieved from http://www.businessinsider.com: http://www.businessinsider.com/netflix-tests-skip-intro-button-to-improve-binge-watching-2017-3

Nelson, L. D., Meyvis, T., & Galak, J. 2009. Enhancing the television-viewing experience through commercial interruptions. *Journal of Consumer Research*, *36*(2), 160-172.

Oxford Dictionary. 2017. Definition of binge-watch in English. Retrieved from oxforddictionaries.com: http://www.oxforddictionaries.com/us/definition/american_english/binge-watch

Schweidel, D. A., & Moe, W. W. 2016. Binge Watching and Advertising. *Journal of Marketing*.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.

TiVo. 2015. *Original Streamed Series Top Binge Viewing Survey for First Time*. Retrieved from pr.tivo.com: http://pr.tivo.com/manual-releases/2015/Original-Streamed-Series-Top-Binge-Viewing-Survey

West, K. 2013. *Unsurprising: Netflix Survey Indicates People Like To Binge-Watch TV*. Retrieved from cinemablend.com: http://www.cinemablend.com/television/Unsurprising-Netflix-Survey-Indicates-People-Like-Binge-Watch-TV-61045.html

Yoganarasimhan, H. 2016. Search personalization using machine learning. *Working Paper.*

Zou, H., & Hastie, T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301-320.