pilot

Audience Relationship Management & ARTIFICIAL INTELLIGENCE

White Paper

Audience Relationship Management & AI

Deriving insights through machine learning

White Paper

Pilot Analytics Inc.

Contents

1 Introduction

In today's world of enterprise-scale artificial intelligence and machine learning, businesses face increasing pressure to integrate data analytics into their regular decision-making processes, and to constantly iterate on the valuable insights provided by increasingly large and multi-dimensional customer datasets.

Yet, as the quality and quantity of data have consistently hewn to a steady upward trajectory, the accessibility and availability of practical analytical solutions has remained stagnant. Especially in the entertainment industry, companies face the difficult task of processing massive amounts of qualitative and quantitative customer data to develop sophisticated content recommendation engines and forecasting algorithms.

With so much on the line — financially, creatively, and temporally — traditional methods such as comps analyses and logistic regressions often prove insufficient in today's competitive data-driven landscape. Technology monopolists such as Netflix and Amazon possess talented data science teams and have already established vertically integrated production and distribution pipelines in the entertainment industry with massive returns to scale.

For existing film industry stakeholders, it is insufficient to merely play catch-up with technology companies built on the ethos of exponential advancement dictated by Moore's Law. In order to remain competitive, it is imperative that entertainment companies take immediate action to ameliorate their existing dearth of data analytics talent. By thoughtfully and deliberately integrating artificial intelligence and machine learning into day-to-day decision-making processes, entertainment companies can establish a virtuous cycle of audience engagement driven primarily by predictive targeted advertising and granular box office forecasting.

2

Audience Relationship Management

2.1 Understanding Audiences

According to the MPAA, almost 250 million North Americans purchased a movie ticket at least once during 2016.[1] Yet only a small fraction of the moviegoing population — less than 10% — is a member of a theater loyalty program. As a result, movie theaters — and every other upstream stakeholder from advertisers to producers — are left in the dark about who is actually watching their films.

Every year, millions of dollars are spent on untargeted advertisements that ultimately never reach their target audience; even worse, industry stakeholders have very few tools to gauge the overall satisfaction of audience members outside of preliminary focus groups and online ratings such as Rotten Tomatoes scores. With ticket prices rising, disappointed audience members leaving the multiplex may opt to reapportion their limited entertainment time and budget toward on-demand streaming content instead.

2.2 Recommender Systems

It has been previously demonstrated by researchers competing in the Netflix Prize competition that collaborative filtering is an effective approach to developing more accurate recommendation systems, specifically within the context of a film dataset. Specifically, a blend of Restricted Boltzman Machines (RBMs) and matrix factorization was shown to be highly performant in a production environment.[2]

[1] MPAA Theatrical Market Statistics Report (2016)
[2] The BellKor Solution to the Netflix Grand Prize: Koren, Bell, and Volinksy (2009)

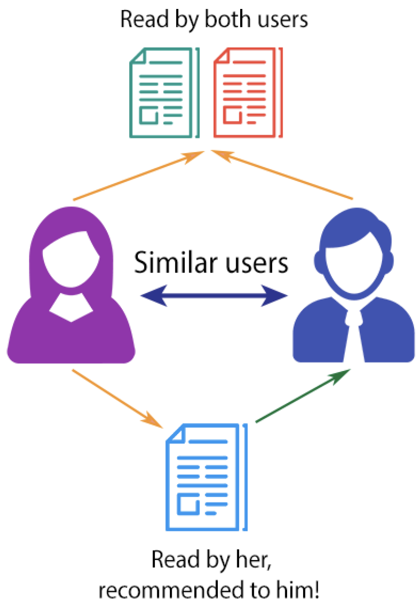
Naively, a collaborative filtering approach relies on existing user data to compute the similarity between individual users or movies. In this example, a user u rates some movie m as an aggregate calculation of similar users' ratings of the same movie:

Example 2.1 (Naive collaborative filtering)

$$r_{u,m} = \text{aggr}_{u' \in U}\, r_{u',m} \tag{2.1}$$

In layman's terms, we can consider two hypothetical individuals A and B who both enjoy a particular film f_1; given that this is the case, we might assume that A and B also share more similar preferences for some subset of additional films within $f_2..f_n$ than with some disparate individual C.

Figure 2.1: An example of collaborative filtering.



2.3 The Data Problem & the Social Media Solution

Unfortunately, it is often difficult to accurately infer user preferences from traditional datasets, which typically only display anonymous telemetry at coarse granularities with oblique actionability. Furthermore, many entertainment companies have struggled to unify multiple datasets into a single set of cohesive user stories across various platforms, devices, and websites.

However, existing research by Kosinski, Stillwell, and Graepel[3] suggests that an individual’s passive social media footprint can be predictive of personal attributes (both phenotypic and psychological) as well as cultural and political beliefs.

Accordingly, entertainment companies must be acutely aware of audience privacy — especially across territories with different data regulations — and serve as de facto fiduciary stewards of their audience’s data. The future of entertainment analytics lies in the eventual cross-device and cross-platform unification of audience data and preference recommendations through a few key social media platforms: Facebook, Instagram, Twitter, Snapchat, and YouTube.

2.4 Audience Relationship Management via Sentinel

Pilot has created a novel AI tool called Sentinel based on novel research conducted by our team at Harvard University. Sentinel leverages sophisticated machine learning techniques to perform collaborative filtering on a proprietary dataset of film viewership.

Put simply, Sentinel clusters audience viewership into so-called "neighborhoods of taste," at which point it then links individual viewers to their specific social media footprint in an anonymized way.

Through collaborative filtering, Sentinel allows stakeholders to determine which demographic cross-sections of an audience comprise the optimal set of viewers for a particular film. It goes one step further by also analyzing the audience’s longitudinal social media footprint for socioeconomic, political, and cultural indicators that reveal specific brand preferences for unique cross-marketing opportunities.

For instance, Sentinel can identify whether a film’s audience is more likely to read a specific news publication relative to a baseline likelihood across the full audience population of moviegoers. Although Sentinel can be robustly applied to large datasets for blockbuster tentpoles, it can just as effectively be used for audience relationship management for smaller films.

[3] Private traits and attributes are predictable from digital records of human behavior: Kosinski, Stillwell, and Graepel (2013)

Example 2.2 (Case Study: *Tangerine*)
The film *Tangerine* (2015) was a Sundance critical darling produced on a $100,000 budget and shot on an iPhone 5S smartphone. It featured multiple transgender people of color in lead roles and received a limited theatrical release that peaked at 44 theaters before finding a second life on Netflix.

Figure 2.2 displays the results of a basic audience analysis for *Tangerine* powered by SENTINEL. We can identify an arbitrary sample of 10,000 audience members who are most likely to enjoy this film based on their past viewing habits, and subsequently observe the general categorical interests of this sample relative to the baseline of the full audience population.

On the following page, Figure 2.3 dives a step deeper into specific categories of interest, which can be further distilled to specific social media pages relating to a single brand, celebrity, or hobby. At this level of granularity, a film distributor or exhibitor can identify brands, publications, and celebrities for directly actionable targeted digital advertising campaigns.

Figure 2.2: General predicted viewer preferences for *Tangerine*

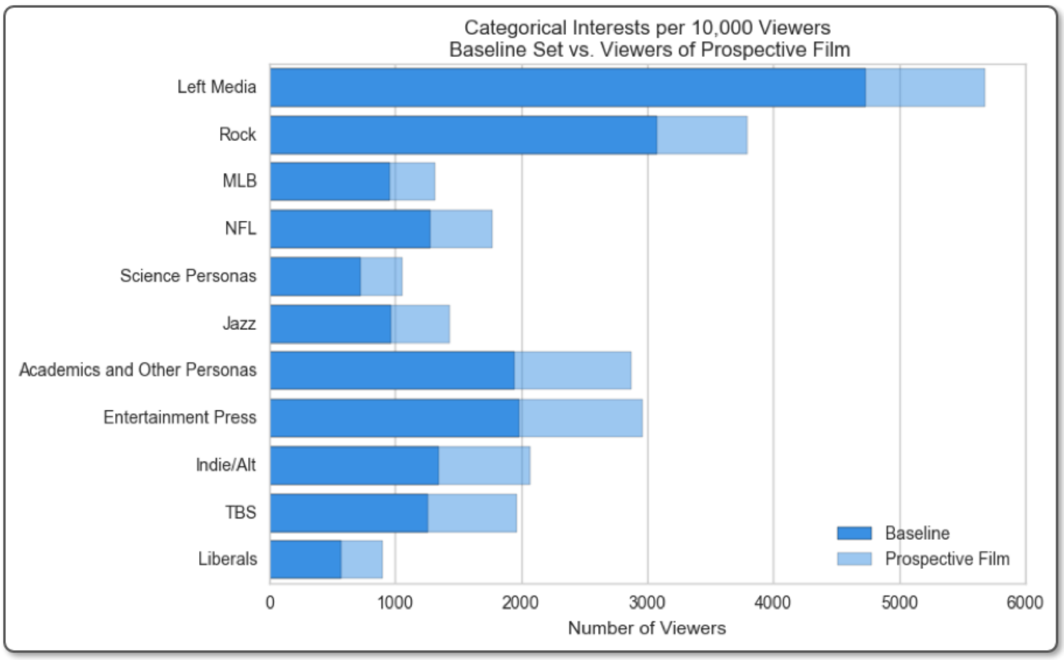
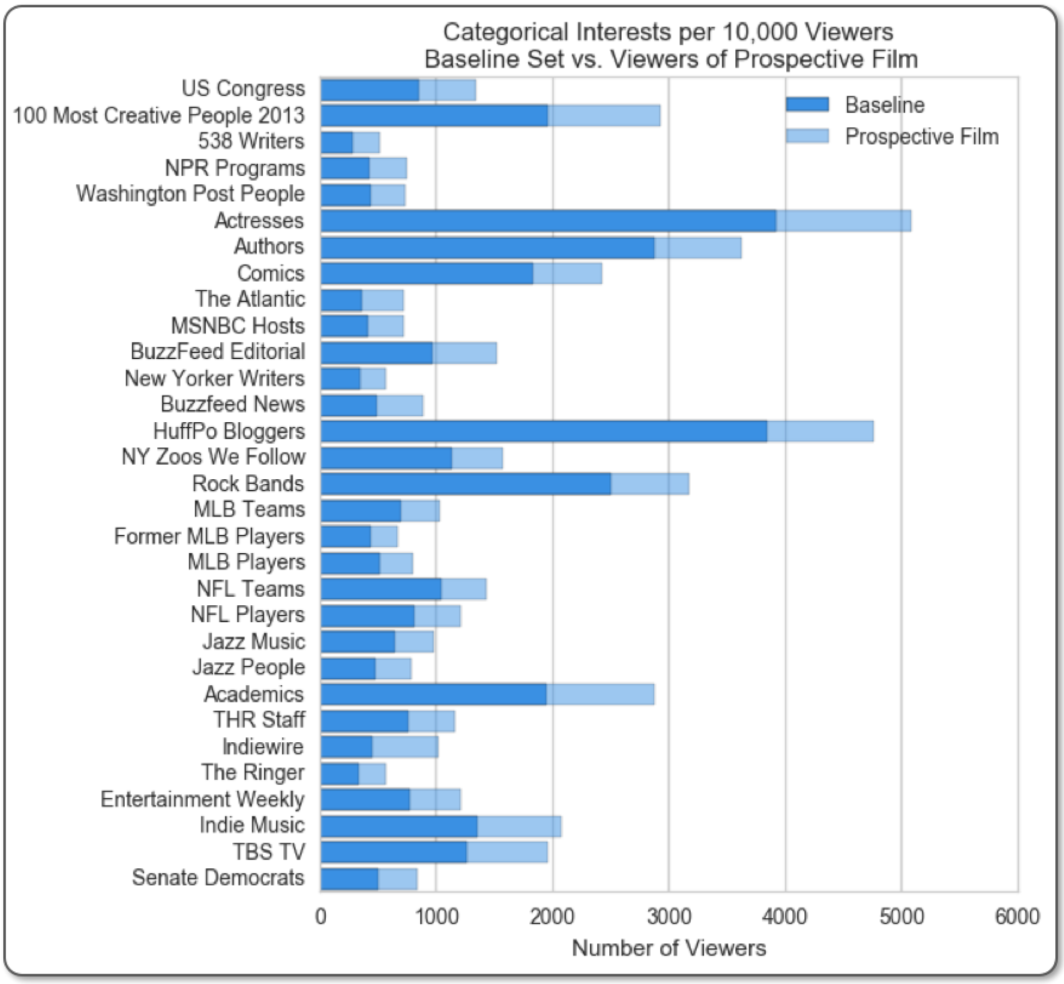


Figure 2.3: Specific predicted viewer preferences for *Tangerine*



When combined with box office forecasting, SENTINEL is a powerful tool for audience relationship management. By identifying the specific interests of each audience member, we can help provide them with an optimal, directly catered entertainment experience.

3

Box Office Forecasting

3.1 The Prophet Methodology

Pilot's suite of AI-based box office forecasting algorithms is called Prophet. For both blockbusters and independent films, Prophet is upwards of 70% accurate in forecasting opening weekend box office from 6-18 months before release, and upwards of 80% accurate once a film's first trailer is released. Forecasts are also available for domestic week-by-week and lifetime box office as well international opening weekend box office for over 50 countries. To arrive at its predictions, Prophet disassembles basic variables such as cast, director(s), writer(s), producers, release date, genere, and plot information into hundreds of secondary features such as network-based variables, markers of seasonality, and country-specific economic indicators.

For example, some key variables estimate the influence of a producer or director along with the cast and writers, while others act as important indicators of how tightly connected cast members are to each other. Consumer spending indices, economic forecasts, and other indicators of consumption are also introduced to the model to adjust individual predictions to the broader economic climate at the time of a film's release. A group of final secondary variables are then chosen through well-established variable selection methods more sophisticated than simple correlations, optimizing for generalizability and assigning more recent movies a larger weight.

Quantifying prior connections between a film's cast and crew as a network plays heavily into the predictions made by Prophet. In one analysis, a network graph of 104 actor nodes displayed 322 edges, each representing collaborations just over the last two years — a level of complexity already far beyond manual interpretation. Over the past several decades, tens of thousands of actors and crew members have generated hundreds of thousands of pairwise collaborations, which contain a wealth of information useful for statistical inference.

Two basic metrics that prove useful in a graph theoretical model of film influence are degree centrality (Equation 3.1) and closeness centrality (Equation 3.2). The former is simply defined as the number of edges connected to a node, while the latter is defined as the average length of the shortest path between a specific node and all other nodes. Both measurements provide critical insight into a particular actor or director's influence in the film industry, but the calculations required to manually compute these values for even a small subset of individuals would be almost intractable.

Example 3.1 (Centrality in a film network)

$$C_D(v) = \deg(v) \tag{3.1}$$

$$C(x) = \frac{1}{\sum_y d(y,x)} \tag{3.2}$$

3.2 Beating the Comps

To its practitioners, film box office forecasting is often seen as more of an art than a science. Indeed, even analysts within the same firm may often disagree on how to best construct a model for a particular film, and how to appropriately weight the plethora of data that is available across historical comps, social media, and other sources.

Pilot bucks the industry trend of primarily relying on a small set of comparable films to calculate revenue forecasts. Our novel approach applies a robust machine learning model trained on a historical dataset of over 4,000 films spanning almost three decades. Since the model is trained on this full dataset, it is as if the full dataset of films serves as the comparables for any given prediction.

Furthermore, our models optimize external statistical validity by only referencing a subset of the overall dataset relative to the theatrical release date of a film. In other words, the model only utilizes information that was available at the time of a movie's release, and doesn't rely on "future" data when making historical predictions.

A unique feature of our forecasts is that each prediction is delivered as a confidence interval. Using a Markov chain Monte Carlo method, the model generates a 95% confidence interval that provides additional actionable insights. For instance, confidence intervals for two different films may have very small or large variances that reflect the model's differing amounts of certainty in each of its predictions. Additionally, because the lower, mean, and upper figures of the confidence interval are not necessarily representative of a normal distribution, we can observe from the actual distribution whether a specific model's predictions tend to skew toward the lower or upper bounds, signifying potential under- or over-performance at the box office.

Figure 3.1: Opening box office prediction for *The Witch*

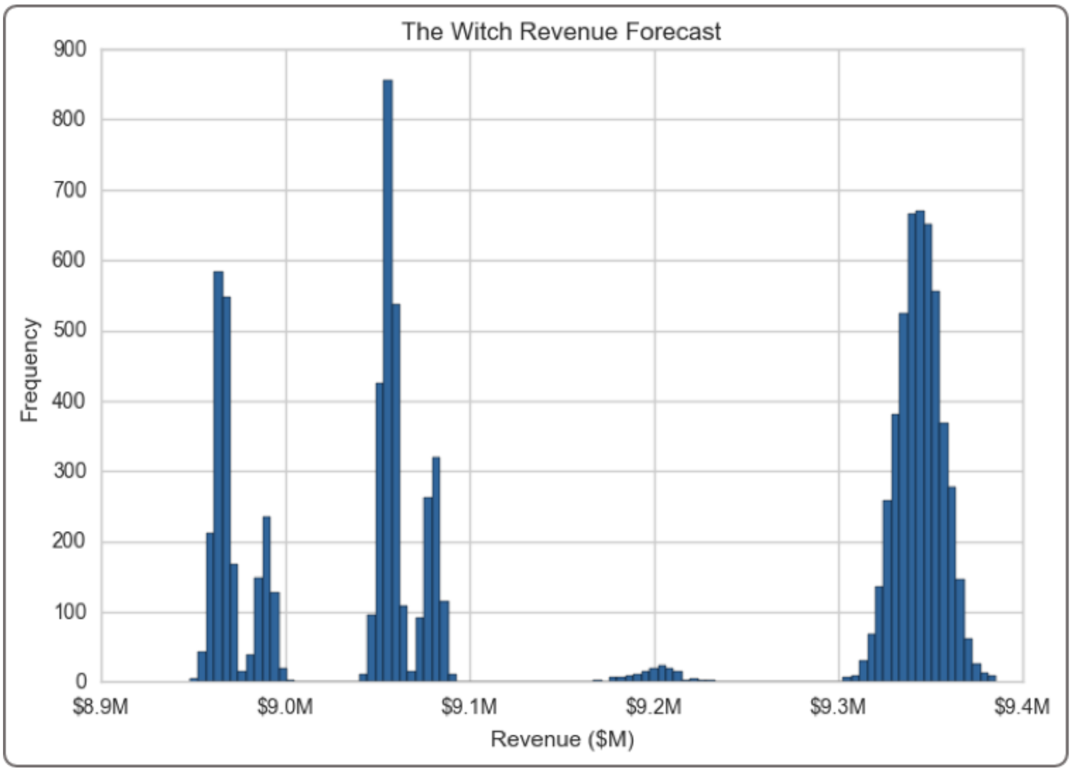


Figure 3.1 displays a sample box office forecast for the opening weekend of A24's *The Witch*, which debuted to a gross of $8.8M. Although the distribution of predictions is multimodal, we can observe from the frequencies that the model believed a prediction of $9M was more likely than $9.5M.

4

Conclusion

As companies seek to establish deeper relationships with their consumers, it is imperative that the film industry leverage artificial intelligence and machine learning to drive more effective data analytics that promote the audience experience.

By pulling on the two levers of targeted advertising and box office forecasting, entertainment companies can drive greater theatrical attendance and derive a better understanding of viewership habits. There are unbounded opportunities in the new paradigm of data-driven entertainment production — it is simply up to us, as innovators, to begin a new virtuous cycle of audience engagement.

Pilot Analytics Inc.
Inquiries: info@pilotmovies.com
Website: https://www.pilotmovies.com
Phone: +1 (323) 208-9016