Advancing predictive content analysis: a natural language processing and machine learning approach to television script data

Anthony Palomba[1]

Abstract

This study introduces a predictive framework for estimating television episode viewership using machine learning and natural language processing applied to over 25,000 TV scripts. By analyzing linguistic and emotional features embedded in dialogue, the research identifies content patterns linked to audience viewership. Multiple regression models, including OLS, Lasso, Ridge, Elastic Net, Gradient Boosting, and XGBoost, are trained to forecast next-episode viewership, explaining up to 50% of variance at the genre level and 41% at the series level. These findings suggest that early-stage script analysis can offer actionable insights for media development and marketing teams. Rather than viewing scripts solely as creative artifacts, this research highlights their potential as data assets for content strategy, allowing for more informed decisions in greenlighting, promotion, and brand alignment.

Keywords Natural language processing · Content analytics · Seriality and engagement theory · TV scripts · Peak-end theory

Introduction

Recent advancements in machine learning and artificial intelligence have facilitated new avenues for content analysis, an established field within communication research that is now expanding into computational and large-scale automated methods. Despite ongoing improvements in recommendation systems, challenges persist with the cold start problem (Yuan & Hernandez 2023), which refers to the difficulty of making accurate content recommendations for new users or new items due to the lack of prior interaction data. As TV series adopt shorter seasons and episodes, crafting rich, character-driven arcs becomes more challenging, though strong narratives remain possible, as seen in films and short stories. Innovation remains limited, especially in proactive content analytics that predict audience resonance. The field still relies on lagging indicators like ratings, making it harder to assess how structured dialogue, used to build plot, tension, and narrative devices, impacts viewers (McKee 1997).

Content analytics is an emerging field spanning media, operations, marketing, and other domains. The study of unstructured data, including text, images, and other creative products, has gained attention from both scholars and practitioners (Balducci & Marinova 2018). With the advent of large language models, new opportunities are arising that could reshape scriptwriting and subsequent production in professional spaces (Luchen & Zhongwei 2023). For example, Eliashberg et al., (2014) demonstrated that using a kernel-based approach allows for predicting the revenue potential of scripts at the green-lighting stage. Despite these advances, little is known about which specific dialogue structures predict episode-to-episode viewer retention.

This study builds on prior efforts to apply content analytics to media narratives, shifting the focus from box office revenue, an imperfect proxy for audience satisfaction, to the relational nature of TV viewing, where sustained return offers a clearer signal of engagement. TV series, by their episodic structure, provide a unique opportunity to examine how dialogue features predict continued viewership. Previous research has laid a foundation for the application of STEM methodologies to narrative media, particularly in the context of movie and TV scripts. For instance, large language models have been used to generate movie summaries and predict genres, demonstrating substantial variation

✉ Anthony Palomba
apalomba87@gmail.com

[1] Darden School of Business, University of Virginia, Charlottesville, VA, USA

in performance across genres (Raj et al. 2024). Similarly, script generation using term frequency-inverse document frequency (TF-IDF) and long short-term memory (LSTM) networks have been explored to understand neural network-based script generation (Dharaniya et al. 2023) along with how textual analysis and natural language processing can be used to examine consumer sentiment and emotions, providing valuable empirical troves of information (Xu et al. 2021; Petrescu & Krishen 2023; Gross & Desveaud 2024). Efforts have also been taken to examine how consumers engage in language across Twitter medium platforms (Murthy et al. 2015) and nuanced media information processing (Geiger & Newhagen 1993).

This study proposes an integrated theoretical framework: peak-end theory explains how emotional high points and endings drive audience retention; seriality and engagement theory describe how dialogue sustains audience investment across episodes; and natural language processing enables systematic measurement of emotional intensity, word count, and linguistic complexity at scale. Together, this study offers a novel methodological framework to advance measurement practices within communication research. Specifically, it integrates natural language processing, ensemble machine learning, and econometric modeling to quantify and predict narrative viewership at scale.

SVOD industry

The rise of machine learning and AI is challenging traditional roles in television writing and reshaping how content success is measured by media professionals. Amid the contentious strikes across the PGA, DGA, and WGA, discussions on the use of natural language processing and the impact of content analytics on Hollywood were central to collective bargaining negotiations (Scheiber & Koblin 2023). Traditionally, audiences have long served as arbiters of content success, helping Hollywood executives gauge performance. This model was essential for mass audiences but contrasts sharply with the current atomization of viewership across platforms and consumer segments (Kim 2018). Audience viewership has long been the gold standard for content success, but as a lagging indicator, it limits timely insights. To better guide advertisers and creatives, we need new frameworks that make content development more data-informed and forecastable.

Peak-end theory

One approach to understanding how people evaluate experiences is by focusing on the most emotionally intense and final moments. Research suggests that individuals tend to judge an experience based on how they felt at its emotional peak(s) and at its conclusion, rather than its overall duration (Fredrickson & Kahneman 1993). Although peak-end theory has seen limited application in media research, recent scholarship highlights its promise (Horowitz et al. 2024). Shehu et al., (2016) found that the beginning, peak, and end of a video strongly influence viral sharing, describing a "rollercoaster effect" in which intense emotional highs and satisfying conclusions outweigh weaker segments. In fact, high-arousal emotions like anger tend to trigger a preference for action, while low-arousal emotions like sadness may lead to withdrawal or inaction (Rucker & Petty 2004).

This expectation is further reinforced by consumer psychology research, which shows that retrospective evaluations are disproportionately shaped by peak moments and how experiences end (Kahneman et al., 1993; Redelmeier & Kahneman 1996). For example, patients recall the most painful and final sensations of medical procedures more vividly than their overall duration. In advertising, Baumgartner et al., (1997) found that consumers prefer advertisements that end on a highly positive scene, and that emotionally steep build-ups toward resilient peaks improve brand recall and liking. van Laer et al. (2014) demonstrate that emotionally rich storytelling boosts both cognitive engagement and behavioral responses, reinforcing the theoretical link between emotional narrative arcs and audience retention in serialized television.

In this study, "peak" moments are operationalized as emotionally intense dialogue identified through LIWC features such as Anger, Fear, or Joy in act 2 (e.g., Anger_2, Fear_2, Joy_2) (rising tension), while "end" moments are represented by similar features in act 3 (resolution). For example, a spike in Anger_2 or Joy_3 in a script segment is treated as a narrative high or emotional closing note. This mapping enables scalable measurement of narrative arcs aligned with peak-end theory, helping to explain patterns of episodic retention across series.

Seriality and engagement theory

Seriality refers to storytelling that unfolds across multiple sequential installments, with narratives evolving to sustain suspense and continuity. For seriality to be effective, stories must maintain internal coherence and plausibility within their established rules and constraints (Hall 2003). Engagement theory further explains how audiences become deeply involved through sustained emotional, cognitive, and behavioral participation (Green & Brock 2000; Green & Clark 2013).

TV is a writer's medium; dialogue is not merely conversation but serves as an essential mechanism to propel the plot, create emotional arcs, and shape character development. According to McKee (1997), well-crafted dialogue must

be economical and structured to generate empathy, tension, and resolution. McKee also warns that repeated exposure to formulaic structures may lead to diminishing returns, underscoring the need for variety and pacing (McKee 1997). For this study, TV is defined as scripted episodic content across traditional broadcast and streaming platforms (e.g., Netflix, Prime Video).

Recent research has expanded understanding of seriality by examining the role of cliffhangers in episodic storytelling. Cliffhangers have been shown to heighten enjoyment, arousal, and intention to continue watching (Wirz et al. 2023; Schibler et al. 2024). Lu et al. (2025) found that unresolved narrative tension versus closure affects subsequent audience behaviors. Narrative engagement should engender consumer pleasure, based on either positive or negative emotions (Oliver 2003; Nabi et al. 2006).

This study applies seriality and engagement theory by analyzing dialogue-based markers of continuity, immersion, and cognitive complexity, specifically, words per sentence, cognitive processing, and a composite engagement variable. Research on over 400,000 books shows that word use peaks mid-story, when cognitive tension is often highest (Boyd et al. 2020). Similarly, features like word count and sentence complexity reflect honesty, engagement, and cognitive effort (Tausczik & Pennebaker 2010). In serialized storytelling, higher linguistic complexity signals deeper narrative involvement and aligns with the mechanisms driving sustained viewer engagement.

Dialogue and story arcs

Dialogue drives plot progression and shapes the viewer experience, serving as a core vehicle for seriality and engagement in TV storytelling. Storytelling transmits information, influences behavior, and fosters social bonds (Bietti et al. 2018), typically through tension-building structures like the three-act framework. The basic narrative framework, which includes exposition, climax, and resolution narrative arc, appears in various formats, such as Supreme Court opinions, novels, and films, with emotional arcs being especially important (Boyd et al. 2020; Dale et al. 2023). TV scripts often follow six familiar story shapes (e.g., rise, fall, fall-rise), identified using NLP by Reagan et al. (2016) and validated by Del Vecchio et al. (2021), who linked these arcs to audience satisfaction using econometric models. These patterns mirror the repetitive structures of prime-time TV, which help audiences recognize and retain narrative expectations (Selnow 1986). McKee (1997) and Srivastava et al., (2023) connect dialogue to narrative momentum via Freytag's pyramid and the three-act structure.

Dialogue scaffolds exposition, conflict, and resolution while shaping narrative worlds. Though its prominence varies by genre, it remains central to clarity and pacing in most scripted TV. However, gaps remain in linking dialogue features to audience behavior. Metrics like word count, sentence length, and emotional conflict have not been systematically tied to viewership. Emotionally charged dialogue fuels tension and retention (McKee 1997; Green & Brock 2000), while linguistic complexity enhances immersion (Boyd et al. 2020; Srivastava et al. 2023). This study analyzes dialogue both within TV series and across genre-level trends.

To clarify the theoretical basis of our hypotheses and enhance construct validity, we explicitly map core concepts to the linguistic features used. Peak-end theory suggests evaluations are shaped by emotional highs and endings; we operationalize this through LIWC-derived features from Act 2 (rising tension) and Act 3 (resolution). High-arousal emotions, whether positive or negative, are more viral and action-inducing (Berger & Milkman 2012). Anger signals agency, while joy invites attention, both linked to control and responsibility (Smith & Ellsworth 1985). Parasocial bonds form when characters display emotional competence and control, especially in concluding scenes (Cohen 2001), increasing admiration and ongoing engagement. Tamborini et al. (2010) argue that media is more rewarding when it fulfills needs for competence, autonomy, and relatedness. Characters with clear motivations and obstacles (McKee 1997) enhance viewer identification, supporting the use of Clout_3 (degree of speaker confidence and social dominance) as a proxy for emotionally resonant endings.

Seriality and engagement theory are captured via linguistic complexity and continuity, key for retention in long-form storytelling (Wirz et al. 2023). These are operationalized through Words per Sentence, Cognitive Process (Cogproc), and composite Engagement scores, reflecting cognitive effort and immersion across acts. Such features encourage reflection and narrative clarity (Green & Brock 2000; Graesser et al. 2014). High scores in Anger_2, Clout_3, or Joy_3 thus indicate peak intensity and meaningful resolution.

Given these literature gaps and clarifications, the present study poses the following hypotheses and research question explicitly leveraging NLP methodologies:

H1 *The level of conflict and tension*, *as measured by NLP-derived dialogue features in act two*, *will strongly predict viewership in millions for the next episode*.

While H1 draws on peak-end theory, H2 and H3 examine observable dimensions of viewership and seriality through narrative structure and linguistic complexity (Boyd et al. 2020; McKee 1997; Mittell 2004; Srivastava et al. 2023). These hypotheses reflect longstanding assumptions that features like word count, sentence complexity, and genre conventions influence audience retention. Prior work shows that grammar, cohesion, and syntactic simplicity affect

cognitive processing and narrative transportation (Green & Brock 2000; Graesser et al. 2014), while narrative richness enhances immersion (Boyd et al. 2020). Genre also shapes expectations for pacing, tone, and character arcs, meaning dialogue features predictive in one genre may be less so in another (McKee 1997; Mittell 2004). This motivates our comparison of individual series versus genre-level patterns. Though viewers do not read scripts, they process dialogue actively, making these linguistic signals crucial for understanding retention. H2 and H3 thus empirically extend narrative theory using scalable, language-based predictors to test how dialogue patterns influence viewership across structural levels.

H2 *Linguistic complexity*, *measured by features like words per sentence and word count across Acts* 1–3, *is expected to significantly correlate with subsequent episode viewership*, *as these elements shape immersion and interpretive effort.*

H3 *The predictive strength of dialogue features is expected to differ between individual series and genre-level aggregates*, *as genre schemas may moderate audience responses to linguistic and emotional cues.*

While the three hypotheses address specific theoretical relationships between dialogue features and viewership, the research question offers a broader lens to evaluate the overall predictive utility of the model. RQ1 complements the hypothesis testing by exploring how well NLP-derived features, segmented by narrative act, can predict viewership outcomes across both individual series and broader genre groupings. This question helps validate the extent to which emotional and linguistic structures contribute to forecasting performance at multiple levels of abstraction.

RQ1 *How effectively do NLP-derived dialogue features from specific episode acts predict subsequent viewership across TV series and TV genres-level groupings?*

Methods

Data preprocessing

The dataset includes TV script text and episode-level viewership across a broad range of series and genres. The unit of analysis is the episode, with dialogue features extracted via natural language processing (NLP). Data were collected in stages. Scripts were scraped from Open Subtitles (2025), chosen for its wide coverage, user engagement, and ratings. An API enabled extraction for 228 TV series, selected from the top 30 by average rating each year since 1984, contingent on script availability. If a show appeared in the top 30 in any year, it was included. On average, 83% of scripts

were obtained per series. Nielsen ratings were sourced from Wikipedia and verified with official data where available. Additional metadata (e.g., network, season/episode count) came from IMDb. The final dataset includes 25,257 episodes ($N = 25,257$).

Appendix 1 outlines the 41 dialogue-based NLP features extracted across Acts 1, 2, and 3, totaling over three million observations ($N = 3,106,611$). These features were derived using the LIWC lexicon, which categorizes words into psychological, emotional, and cognitive domains. LIWC was selected due to its proven effectiveness in analyzing media content, including studies on movie violence (Martinez et al. 2019), character traits in film (Ramakrishna et al. 2017), and parasocial relationships on Twitch (Leith 2021). It captures core script elements such as emotional tone, cognitive focus, and social dynamics (Tausczik & Pennebaker 2010).

Key variables include Clout, Engaged (a composite of cognitive and emotional engagement), and emotional features such as Anger, Joy, Fear, Sadness, and Disgust. Variables with suffixes _1, _2, and _3 correspond to Acts 1 (exposition), 2 (rising tension), and 3 (climax/resolution), respectively; for instance, Anger_2 captures anger in Act 2, while Clout_3 reflects dominance in Act 3. A full glossary appears in Appendix 1. To mitigate spurious correlations, feature selection was performed using random forest regression, with VIF analysis applied to address multicollinearity and fivefold cross-validation used to ensure model robustness.

To prepare the dataset, the dependent variable of interest was viewership in millions for the next episode, and so this variable was shifted by one episode to create a lag, with the last row of each show being dropped to avoid NaN (Not a Number) values, which represent missing or undefined numerical data in the dataset. Shows with fewer than five episodes after lagging were excluded from the analysis to ensure robust sample sizes. To prepare the data for regression, features irrelevant to the analysis (e.g., show name, season, episode number, and network information) were removed. Only numeric columns were retained for the feature matrix, which was scaled using MinMaxScaler to normalize the range of all features (Géron 2019).

While viewership is shaped by external factors (e.g., marketing, scheduling, platform access), it remains a widely accepted behavioral proxy for content success and sustained audience interest. Research in media and advertising links higher viewership to greater attention, emotional investment, and narrative immersion (Bourdon 2021; Rubin 2009). It also remains the industry standard for evaluating episodic performance, informing renewals, ad rates, and platform strategy (Napoli, 2011). Nelson and Webster (2016) note that metrics like audience size and time spent act as market "currencies," capturing both exposure and behavioral engagement. Though conceptually distinct, tuning duration

and "attention minutes" empirically bridge exposure and engagement by reflecting audience loyalty.

Feature selection and model fitting

Feature selection was conducted using a random forest regressor model, trained on the scaled features to compute feature importance scores. A random forest regressor is a robust ensemble learning method that constructs multiple decision trees and aggregates their predictions to enhance accuracy. This algorithm optimizes each tree split by maximizing the reduction of a specific error metric, such as mean squared error (MSE) or mean absolute error (MAE) in an effort to refine predictive performance. As a decision tree-based approach, it combines bagging and feature selection techniques to improve predictive performance (Breiman 1996; Amit & Geman 1997). Consequently, the features selected at each node are those that most effectively minimize error, and they are ranked based on their contribution to error reduction (Kursa & Rudnicki 2011; Dewi & Chen 2019). Furthermore, when analyzing two hundred and twenty-eight TV series at the series level, random forest proves advantageous in reducing computational complexity, thereby enabling more efficient analyses (Yates & Islam 2021) and feature isolation (Géron 2019).

For each TV series, the ten most important features were retained for further analysis. To account for potential multicollinearity, the variance inflation factor (VIF) was calculated for all selected features. Features with VIF values exceeding the threshold of ten were iteratively removed until all remaining features fell below this threshold (Hair et al. 2018). If fewer than ten features remained, additional features with the lowest VIF or highest importance were reintroduced to ensure a minimum of ten features per regression. For each show, an OLS regression model was fitted using the selected and filtered features.

To examine genres, different genres were aggregated to reduce computational need and enhance predictive power. These included comedy (comprised of sitcoms, comedy drama, animated sitcom, comedy, and comedy horror), drama (composed of medical dramas, serial dramas, family dramas, action dramas, legal dramas, legal thriller, political drama, psychological drama, period drama), crime and thriller (composed of crime drama, police procedural, crime, legal thriller, and psychological thriller), action and adventure (composed of action, action-adventure, superhero, adventure), fantasy and science fiction (composed of fantasy, science fiction, science fantasy, supernatural, supernatural drama), and horror and mystery (composed of horror, mystery, and psychological horror). In certain cases, genres were intentionally included in more than one aggregate group to reflect overlapping thematic and structural elements across categories; this flexible assignment acknowledges the hybrid nature of many modern TV series. Grouping genres into broader aggregate categories, even with some overlap, offers a pragmatic way to capture the complexity of genre relationships and their influence on viewership. TV genres often share thematic and narrative elements, making rigid classifications limiting. Aggregating allows for more flexible analysis and better reflects the multifaceted nature of modern TV series.

To handle this complexity, advanced machine learning models such as XGBoost, Gradient Boosting, and ElasticNet were employed, along with Shapley values for feature importance. These methods offer a robust, scalable alternative to traditional models like OLS and Random Forest. To account for temporal dynamics, viewership was lagged by one episode to predict future performance based on prior dialogue. Initial models also included seasonality (e.g., time of day, day of week). While variables like episode length and season count were strong predictors, they were excluded from final models to isolate the effect of dialogue features, and avoid confounding from production scale or longevity.

Model evaluation

The regression models were evaluated using R^2, adjusted R^2, mean absolute error (MAE), and root mean square error (RMSE). Additionally, fivefold cross-val-idation was applied to each show's dataset, using R^2 and adjusted R^2 as scoring metrics. Cross-validation, a resampling technique, assesses model strength by testing it against holdout samples, ensuring a well-optimized train-test split (Allen 1974; Geisser 1975; Gunasegaran & Cheah 2017). Splitting the test set into five parts allows for more accurate validation of model performance (Géron 2019). All regression results, including R^2, adjusted R^2, MAE, RMSE, and cross-validation scores, were stored in a results list for each series.

From the 228 series analyzed, the top 100 were selected based on cross-validation mean R^2 scores, which test model predictability on unseen data. Here, models trained on 80% of each series were tested on the remaining 20%. Feature selection was then conducted using a random forest regressor to identify the top ten most important dialogue features for each show. Subsequent analyses focused on these top 100 series (Table 1), which included titles like The Cosby Show, Home Improvement, Seinfeld, and Elementary, notable for their strong serialized structures.

Given the negative test adjusted R^2 for some variables, it was clear the model was overfitting to the training data. Its test performance was worse than a simple mean-based model, indicating poor generalization. To address this, elastic net and gradient boosting were used. Elastic net applied flexible regularization by combining L1 and L2 penalties, while gradient boosting captured non-linear patterns

Table 1 Top one hundred CV R² TV series evaluation scores	Title	Test CV mean R²	Number of scripts	Test adjusted R²	Test MAE	Test RMSE
	The Cosby Show	0.85	98	− 0.61	5.60	7.25
	Home improvement	0.80	203	− 0.08	5.65	6.97
	Seinfeld	0.80	165	0.34	3.33	4.45
	Elementary	0.75	153	− 0.26	2.20	3.25
	Roseanne	0.69	225	0.52	3.95	4.90
	In Living Color	0.66	124	− 0.38	3.06	4.00
	Malcolm in the Middle	0.64	150	− 0.13	2.79	3.71
	A Different World	0.62	93	− 0.34	4.44	5.35
	Scrubs	0.61	180	0.12	2.30	3.21
	CSI: Crime Scene Investigation	0.61	334	0.60	3.47	4.36
	Grey's Anatomy	0.60	415	0.44	3.13	4.01
	Just Shoot Me!	0.59	144	− 0.02	2.61	3.48
	The Golden Girls	0.58	96	− 1.24	4.11	4.92
	Lost	0.57	114	− 0.46	2.60	3.40
	Desperate Housewives	0.55	179	− 0.13	4.70	5.49
	Designing Women	0.55	116	− 0.23	3.44	4.18
	NCIS: Los Angeles	0.55	315	0.42	2.26	2.98
	King of the Hill	0.54	254	0.24	2.87	3.59
	House	0.51	175	0.28	2.90	3.62
	Full House	0.50	164	0.00	2.38	3.07
	Law & Order	0.50	487	0.40	2.87	3.55
	Empty Nest	0.50	168	0.72	2.57	3.24
	Family Matters	0.49	95	− 0.51	2.23	2.88
	Bull	0.48	125	0.48	0.94	1.59
	Futurama	0.46	82	− 1.33	2.81	3.45
	Doogie Howser, M.D	0.46	96	− 0.72	2.27	2.88
	The Walking Dead	0.46	177	− 0.07	3.39	3.98
	NYPD Blue	0.46	260	0.53	2.21	2.80
	Two and a Half Men	0.45	261	− 0.02	1.90	2.49
	Frasier	0.44	256	0.24	3.18	3.77
	Game of Thrones	0.42	73	− 2.98	1.72	2.31
	Family Guy	0.42	407	0.05	2.16	2.74
	Mad About You	0.41	156	0.27	2.18	2.75
	Criminal Minds	0.40	306	0.41	1.82	2.39
	CSI: Miami	0.40	232	− 0.01	2.51	3.04
	Without a Trace	0.39	156	0.31	1.74	2.22
	NCIS	0.39	457	0.57	2.12	2.59
	Bones	0.39	244	0.13	1.68	2.14
	The Mentalist	0.38	149	0.08	2.04	2.48
	Everybody Loves Raymond	0.38	208	0.50	1.75	2.19
	The Blacklist	0.36	218	− 0.09	1.88	2.31
	Modern Family	0.36	249	0.23	1.72	2.13
	The Big Bang Theory	0.36	279	0.13	1.99	2.38
	Black-ish	0.36	166	0.24	0.80	1.19
	Person of Interest	0.36	103	− 0.01	1.75	2.13
	The Nanny	0.36	143	0.08	2.41	2.80
	Scorpion	0.35	92	0.38	0.81	1.20
	Ugly Betty	0.35	85	− 0.01	1.71	2.08
	The King of Queens	0.35	203	0.12	1.55	1.91

Table 1 (continued)

Title	Test CV mean R²	Number of scripts	Test adjusted R²	Test MAE	Test RMSE
NCIS: New Orleans	0.35	149	0.44	1.86	2.22
The West Wing	0.34	148	0.12	2.04	2.39
24	0.34	192	0.15	1.29	1.61
Crossing Jordan	0.34	104	0.18	1.12	1.44
Ally McBeal	0.33	111	− 0.08	1.46	1.77
Body of Proof	0.33	42	None	0.82	1.12
Blue Bloods	0.33	272	− 0.09	1.61	1.90
Medium	0.32	128	0.49	1.23	1.53
Mom	0.32	168	0.04	0.91	1.20
Heroes	0.31	77	− 0.29	2.02	2.29
2 Broke Girls	0.31	135	− 0.02	1.25	1.51
Becker	0.31	128	− 0.08	1.89	2.14
JAG	0.31	225	0.32	1.41	1.67
Designated Survivor	0.31	43	None	0.63	0.88
CSI: NY	0.30	196	0.34	1.01	1.25
Hawaii Five-0	0.30	233	− 0.22	0.90	1.13
Mike & Molly	0.30	95	− 0.53	1.12	1.35
Rules of Engagement	0.30	95	− 0.01	1.17	1.39
Cold Case	0.29	154	0.48	1.22	1.44
Las Vegas	0.29	103	0.18	1.02	1.24
The Good Wife	0.29	152	0.57	0.84	1.05
Ghost Whisperer	0.28	107	− 0.38	0.87	1.07
Breaking Bad	0.28	62	− 0.79	0.64	0.85
Scandal	0.28	121	− 0.37	1.34	1.54
Life	0.27	32	None	0.50	0.70
FBI	0.27	101	0.12	0.51	0.70
Deadwood	0.26	24	None	0.53	0.71
Chicago Med	0.26	167	− 0.18	0.64	0.83
Community	0.25	90	− 1.55	0.75	0.93
Shark	0.25	37	None	1.01	1.18
Madam Secretary	0.25	120	0.75	0.86	1.03
30 Rock	0.24	135	0.09	0.74	0.89
Sons of Anarchy	0.24	92	− 0.81	0.52	0.67
Blindspot	0.24	100	0.47	0.87	1.02
Parks and Recreation	0.23	122	0.22	0.41	0.55
S.W.A.T	0.22	126	− 0.07	0.63	0.78
True Blood	0.22	80	− 0.70	0.75	0.90
Life in Pieces	0.21	76	0.23	0.65	0.79
Arrow	0.21	170	0.20	0.54	0.67
Mad Men	0.21	90	− 1.14	0.41	0.53
Young Sheldon	0.21	126	0.83	0.63	0.74
Big Love	0.21	49	None	0.41	0.52
A Million Little Things	0.20	79	− 0.41	0.68	0.79
The Shield	0.20	88	− 0.36	0.57	0.68
Dexter	0.19	96	− 0.30	0.40	0.48
It's Always Sunny in Philadelphia	0.19	146	− 0.71	0.43	0.51
Adventure Time	0.19	248	0.46	0.37	0.44
Archer	0.19	102	0.08	0.28	0.35
Homeland	0.18	96	− 0.53	0.25	0.32

Table 1 (continued)	Title	Test CV mean R²	Number of scripts	Test adjusted R²	Test MAE	Test RMSE
	Broad City	0.18	46	None	0.24	0.27
	Fargo	0.17	51	None	0.18	0.21

effectively (Géron 2019). GridSearchCV optimized alpha for elastic net and tuned learning rate and max depth for gradient boosting, balancing complexity and performance. These models improved average CV mean R^2 to 39% and 40%, respectively, and significantly increased test adjusted R^2, confirming better generalization. For further tuning, the Optuna library was used to automate hyperparameter search through dynamic space construction and pruning strategies, efficiently optimizing model performance (Akiba et al. 2019).

For gradient boosting, the best hyperparameters were 200 estimators, a learning rate of 0.05, max depth of 5, and a subsample ratio of 0.85. This setup balances complexity and efficiency, with the low learning rate ensuring gradual adjustments and the subsample ratio helping prevent overfitting. For Elastic Net, the optimal alpha was 0.1 and L1 ratio 0.6, favoring flexibility and slight Lasso emphasis for feature selection. While Optuna did not yield significant gains, both models demonstrated that dialogue-based features can meaningfully predict subsequent viewership at both the TV series and genre levels, directly addressing RQ1.

Data results

The analysis proceeded in three stages. First, random forest was used for feature selection on NLP dialogue variables. Second, elastic net and gradient boosting models were applied to predict next-episode viewership, with hyperparameters tuned via Optuna. Third, Shapley values and two-stage modeling helped interpret which dialogue features most influenced predictions. Table 2 presents model results by TV series. This variability supports RQ1 and aligns with the theoretical framing of seriality and engagement, suggesting that dialogue structures, especially those tied to continuity or emotional climax, are more predictive in some genres. The elastic net model also tested the linearity of dialogue–viewership relationships, blending lasso's feature selection with ridge's control for multicollinearity (Géron 2019).

Programs like *Young Sheldon* (0.891)—a rare comedy with highly structured dialogue—and *Breaking Bad* (0.875) showed strong predictive performance in the elastic net model, suggesting a strong linear relationship between NLP dialogue features and viewership. These series, along with *Madam Secretary* (0.860), *Scorpion* (0.720), *Blindspot* (0.710), and *Bull* (0.707), featured tight story structures

with consistent pacing, clear act segmentation, and structured dialogue arcs aligned to narrative beats (e.g., conflict, climax, resolution), as captured by act-based NLP features. These shows also followed clear procedural rhythms with well-defined turning points and predictable dialogue patterns (e.g., word count, tension terms). Further down the R^2 ranking, procedurals like *NCIS: New Orleans* (0.658), *CSI* (0.642), and *NYPD Blue* (0.632) also appeared.

In contrast, elastic net models for comedies, often centered on light-hearted humor and character interactions, showed weaker predictive performance due to less structured dialogue arcs. Serialized dramas like *Breaking Bad* (0.875) and *The Good Wife* (0.671) produced higher R^2 values than episodic comedies such as *30 Rock* (0.481), *Frasier* (0.450), *Modern Family* (0.414), and *The Nanny* (0.387), suggesting that evolving narratives with emotional peaks provide more consistent signals for NLP-based prediction. Serialized dramas offer deeper, more coherent dialogue, while episodic formats often lack the continuity needed for similar predictability. Procedurals like *CSI* (0.642) and *Cold Case* (0.604) also performed well, while action-heavy shows like *NCIS: Los Angeles* (0.505) and *24* (0.344) had lower R^2 values, likely due to limited dialogue depth. Long-running series like *Law & Order* (0.453) and *Grey's Anatomy* (0.461) showed moderate predictability, indicating that even amid creative fatigue, their core dialogue themes remain relatively stable.

In addition to elastic net, gradient boosting was applied to capture non-linear relationships in dialogue and narrative structure. This ensemble method iteratively fits predictors to residuals, allowing it to detect higher-order effects across complex story arcs. Series like *Young Sheldon* (0.923), *Madam Secretary* (0.882), *Deadwood* (0.959), *Ugly Betty* (0.749), *Adventure Time* (0.749), and *Black-ish* (0.748) showed strong performance, suggesting that richly layered dialogue helps drive plot predictability. In contrast, hybrid action dramas like *Heroes* (0.386) and *The Mentalist* (0.244) yielded lower R^2 values, likely due to their reliance on nonverbal or visual storytelling elements, which reduce the predictive power of dialogue-based models.

To balance predictive accuracy with interpretability, a two-stage modeling approach was used, integrating machine learning and econometric techniques. First, average predictive scores were computed from Elastic Net and Gradient Boosting outputs, combining Elastic Net's feature selection and multicollinearity control with Gradient Boosting's

Table 2   Elastic Net and Gradient Boosting scores for top one hundred CV R² TV series

TV series	Elastic Net R²	Elastic Net MAE	Elastic Net RMSE	Gradient Boosting Net R²	Gradient Boosting MAE	Gradient Boosting RMSE
Young Sheldon	0.891	0.643	0.783	0.923	0.511	0.659
Breaking Bad	0.875	0.474	0.546	0.535	0.673	1.055
Madam Secretary	0.860	0.870	1.021	0.882	0.882	0.935
Empty Nest	0.810	2.623	3.215	0.687	3.132	4.126
Scorpion	0.720	0.844	1.212	0.726	0.853	1.198
Blindspot	0.710	0.888	1.103	0.814	0.692	0.885
Bull	0.707	1.021	1.557	0.835	0.883	1.168
The Good Wife	0.671	0.905	1.126	0.735	0.735	1.010
NCIS: New Orleans	0.658	1.761	2.143	0.668	1.464	2.112
Ugly Betty	0.650	1.547	2.002	0.749	1.335	1.693
CSI: Crime Scene Investigation	0.642	3.635	4.494	0.687	3.151	4.200
Crossing Jordan	0.634	1.092	1.360	0.559	1.192	1.492
NYPD Blue	0.632	2.217	2.767	0.613	2.133	2.841
Roseanne	0.632	4.004	4.894	0.586	4.279	5.194
Life in Pieces	0.628	0.812	1.027	0.853	0.476	0.644
Rules of Engagement	0.625	0.955	1.275	0.491	1.198	1.485
Everybody Loves Raymond	0.611	1.765	2.208	0.451	2.098	2.624
Cold Case	0.604	1.341	1.537	0.584	1.350	1.574
NCIS	0.600	2.170	2.653	0.640	1.976	2.516
Life	0.596	0.549	0.670	0.699	0.499	0.579
Shark	0.594	1.247	1.378	0.673	1.061	1.236
Without a Trace	0.575	1.711	2.144	0.464	1.669	2.408
A Million Little Things	0.564	0.665	0.758	0.302	0.808	0.959
Adventure Time	0.559	0.369	0.445	0.749	0.271	0.336
Seinfeld	0.553	3.479	4.397	0.411	4.098	5.049
Archer	0.543	0.279	0.346	0.548	0.297	0.344
FBI	0.536	0.583	0.741	0.461	0.573	0.799
CSI: NY	0.532	0.996	1.227	0.450	1.017	1.330
Black-ish	0.514	0.788	1.146	0.748	0.621	0.826
Las Vegas	0.512	1.128	1.354	0.545	1.051	1.308
Person of Interest	0.511	1.682	2.103	0.534	1.582	2.053
NCIS: Los Angeles	0.505	2.277	3.005	0.421	2.269	3.250
Deadwood	0.494	0.500	0.690	0.959	0.183	0.195
30 Rock	0.481	0.715	0.857	0.429	0.703	0.898
House	0.478	2.900	3.675	0.381	3.100	4.000
Big Love	0.469	0.488	0.579	− 0.197	0.446	0.446
Grey's Anatomy	0.461	3.262	4.209	0.459	3.114	4.216
Parks and Recreation	0.459	0.469	0.603	0.655	0.351	0.481
Arrow	0.454	0.533	0.664	0.520	4.803	0.622
Law & Order	0.453	2.797	3.585	0.617	2.352	3.001
Heroes	0.452	2.070	2.595	0.386	2.429	2.746
Frasier	0.450	3.060	3.587	0.264	3.378	4.149
A Different World	0.439	4.652	5.197	0.542	4.131	4.696
Criminal Minds	0.437	1.987	2.550	0.488	1.823	2.433
Ally McBeal	0.429	1.501	1.774	0.370	1.481	1.863
True Blood	0.425	0.748	0.904	0.504	0.602	0.840
The West Wing	0.424	2.055	2.385	0.366	2.037	2.502
Mad About You	0.422	2.448	2.997	0.311	2.260	3.273

Table 2 (continued)

TV series	Elastic Net R²	Elastic Net MAE	Elastic Net RMSE	Gradient Boosting Net R²	Gradient Boosting MAE	Gradient Boosting RMSE
The Shield	0.414	0.576	0.693	0.466	0.560	0.662
Modern Family	0.414	1.669	2.092	0.324	1.869	2.247
Body of Proof	0.413	0.856	1.233	0.303	1.047	1.344
Dexter	0.411	0.406	0.481	0.533	0.384	0.428
JAG	0.403	1.497	1.773	0.529	1.245	1.574
Becker	0.393	1.826	2.077	0.651	1.269	1.574
The Nanny	0.387	2.453	2.853	0.454	2.332	2.692
2 Broke Girls	0.369	1.182	1.519	0.469	1.181	1.393
Scrubs	0.368	2.319	3.218	0.247	2.642	3.515
Mike & Molly	0.361	1.129	1.307	0.250	1.214	1.416
Just Shoot Me!	0.355	2.599	3.452	0.358	2.469	3.446
The Mentalist	0.353	2.080	2.563	0.244	2.297	2.772
24	0.344	1.327	1.656	0.339	1.396	1.663
Mom	0.344	0.911	1.188	0.333	0.942	1.197
Medium	0.329	1.724	2.260	0.634	1.276	1.670
Full House	0.323	2.359	3.047	0.231	2.559	3.249
Bones	0.320	1.672	2.125	0.328	1.631	2.113
The Blacklist	0.319	1.676	2.091	0.381	1.607	1.993
Designing Women	0.313	3.218	4.221	0.186	3.848	4.593
Fargo	0.311	0.232	0.253	0.476	0.175	0.220
The Walking Dead	0.306	3.090	3.803	0.258	3.153	3.931
Homeland	0.301	0.253	0.320	0.155	0.263	0.352
Sons of Anarchy	0.299	0.514	0.631	0.492	0.453	0.537
Designated Survivor	0.298	0.519	0.731	0.223	0.569	0.769
Malcolm in the Middle	0.295	2.687	3.624	0.125	3.023	4.037
The Big Bang Theory	0.281	1.970	2.400	0.244	1.871	2.460
The King of Queens	0.258	1.657	2.022	0.383	1.543	1.844
Ghost Whisperer	0.248	0.922	1.097	-0.047	0.982	1.295
CSI: Miami	0.240	2.417	2.985	0.101	2.636	3.246
Desperate Housewives	0.221	4.624	5.398	-0.025	5.036	6.194
The Cosby Show	0.215	5.960	7.345	0.294	5.443	6.969
King of the Hill	0.210	3.307	4.102	0.550	2.407	3.094
Scandal	0.206	1.422	1.564	0.082	1.340	1.681
Chicago Med	0.200	0.626	0.812	− 0.136	0.813	0.968
Mad Men	0.175	0.416	0.515	0.320	0.363	0.467
Futurama	0.175	2.378	3.354	0.516	1.794	2.569
In Living Color	0.162	3.050	4.081	0.264	3.237	3.825
Lost	0.161	2.736	3.481	0.340	2.267	3.089
Family Matters	0.157	2.428	3.225	0.460	1.957	2.581
Hawaii Five-0	0.146	0.836	1.067	0.150	0.861	1.065
Elementary	0.128	2.190	3.317	0.058	2.341	3.448
Blue Bloods	0.119	1.604	1.894	0.053	1.605	1.964
Two and a Half Men	0.110	1.980	2.592	0.061	1.909	2.663
Family Guy	0.068	2.251	2.909	0.419	1.786	2.297
Home Improvement	0.067	6.085	7.466	0.044	5.944	7.560
The Golden Girls	0.049	3.837	4.805	0.181	3.891	4.459
Doogie Howser, M.D	0.017	2.456	3.269	0.189	2.354	2.968
S.W.A.T	0.014	0.840	0.974	− 0.069	0.843	1.014

Table 2 (continued)

TV series	Elastic Net R²	Elastic Net MAE	Elastic Net RMSE	Gradient Boosting Net R²	Gradient Boosting MAE	Gradient Boosting RMSE
Broad City	− 0.015	0.229	0.254	− 0.020	0.200	0.254
It's Always Sunny in Philadelphia	− 0.046	0.428	0.500	0.016	0.410	0.485
Community	− 0.101	0.782	0.953	− 0.135	0.814	0.968
Game of Thrones	− 0.349	1.992	2.514	− 0.269	2.098	2.439

ability to capture non-linear relationships. This hybrid model identified top-performing TV shows based on script features. Second, Ordinary Least Squares (OLS) regression was run on these top-ranked shows to examine which dialogue features specifically drove viewership. Relying on Elastic Net for selection and OLS for interpretation ensured statistical rigor and theoretical clarity.

The top fifty TV series were ranked based on their average R² scores, with particular dialogue features outlined in Table 3 and natural language processing terms detailed in Appendix 1. Across these series, features such as word count, words per sentence (WPS), language complexity (Dic), and psychological involvement or emotional engagement (Engaged) consistently appeared as top predic-tors as illustrated in TV series like 2 Broke Girls (WPS_1, $\beta = − 2.07$, $p < 0.05$; Six Letter Words_3, $\beta = − 3.35$, $p < 0.001$), 30 Rock (WPS_1, $\beta = − 1.84$, $p < 0.05$; WPS_2, $\beta = − 3.72$, $p < 0.01$), Adventure Time (WPS_1, $\beta = − 2.01$, $p < 0.05$). Notably, dialogue features from later episode seg-ments (act three) frequently appeared as important predictors in several series. This is evident across TV series such as Blindspot (Word Count_3, $\beta = -2.49$, $p < 0.05$; Achieve_3, $\beta = − 1.95$, $p < 0.05$), Dexter (WPS_3, $\beta = − 1.24$, $p < 0.001$), Grey's Anatomy (Engaged_3, $\beta = − 7.57$, $p < 0.001$), JAG (Clout_3, $\beta = 3.12$, $p < 0.01$), and Law and Order (WPS_3, $\beta = 9.56$, $p < 0.001$). This provides supporting evidence for H2.

Emotions are also remarkably compelling, as there are numerous instances of fear, joy, sadness, and anger. Fear is a particularly powerful feature, as seen across procedurals like *CSI: Cold Case* (Fear_3, $\beta = 2.20$, $p < 0.05$), and *Rules of Engagement* (Fear_1, $\beta = − 3.46$, $p < 0.001$). Sadness is a dialogue feature that appears to cut across dramas and come-dies, appearing in TV series like *The Good Wife* (Sadness_2, $\beta = 1.20$, $p < 0.05$) and *Life in Pieces* (Sadness_2, $\beta = 1.85$, $p < 0.05$). Joy was a predictive feature for *Criminal Minds* (Joy_3, $\beta = − 5.19$, $p < 0.001$), *The Shield* (Joy_2, $\beta = 0.98$, $p < 0.05$), and *CSI:NY* (Joy_1, $\beta = − 2.71$, $p < 0.01$).

The engagement feature appears in high performing TV series, which suggests that character engagement in a plot is remarkably critical to buy in for audience members. Particularly, being able to have dialogue that is structured around this is exceptionally pertinent. This was found to be

particularly predictive across procedurals such as *Criminal Minds* (Engaged_3, $\beta = − 3.59$, $p < 0.05$), *NCIS* (Engaged_2, $\beta = − 3.33$, $p < 0.01$), *NCIS: New Orleans* (Engaged_2, $\beta = − 4.92$, $p < 0.001$), and *NYPD Blue* (Engaged_1, $\beta = 3.72$, $p < 0.05$) along with comedies like *Archer* (Engaged_2, $\beta = − 0.39$, $p < 0.01$), *Black-ish* (Engaged_2, $\beta = − 1.43$, $p < 0.05$), and *2 Broke Girls* (Engaged_3, $\beta = 1.79$, $p < 0.05$). Another intense emotion, disgust, also appears across numerous TV series. This includes *Parks and Recreation* (Disgust_3, $\beta = 0.88$, $p < 0.05$), *House* (Disgust_1, $\beta = 5.95$, $p < 0.01$), *Law & Order* (Disgust_2, $\beta = 4.63$, $p < 0.001$), and *True Blood* (Disgust_2, $\beta = − 1.82$, $p < 0.01$).

For highly complex TV series, the amount of cognitive processing and analytical thinking tends to be remarkably critical in TV series elucidating strong and critical peak moments for TV series. This dialogue feature is particu-larly prevalent among comedies such as *Adventure Time* (Cogproc_2, $\beta = − 0.71$, $p < 0.05$), *A Million Little Things* (Analytic_2, $\beta = − 1.43$, $p < 0.01$), *Archer* (Cogproc_3, $\beta = − 0.35$, $p < 0.05$), and *CSI: Crime Scene Investigation* (Analytic_3, $\beta = 6.62$, $p < 0.001$). Language complexity (Dic) was a key predictive feature across TV series such as *30 Rock* (Dic_2, $\beta = 1.07$, $p < 0.05$), *Adventure* Time (Dic_1, $\beta = 1.45$, $p < 0.001$), *Empty Nest* (Dic_1, $\beta = − 8.93$, $p < 0.001$), and *Seinfeld* (Dic_3, $\beta = − 18.78$, $p < 0.001$). Across dialogue features that are related to emotions, this provides supporting evidence for H1.

As illustrated in Table 4, advanced robustness models were run across aggregate genres. As illustrated, the dataset provides a robust sample of TV episodes, particularly for comedy (9,140), crime (5,549), drama (3,971), action and adventure (1,816), and fantasy and science fiction (1,326), along with a smaller sample for horror and mystery (388). As illustrated in the table, the XGBoost performance for the top ten features in each model performed remarkably well, scoring between 44.43 and 67% predictive power.

Shapley values were selected to report feature contribu-tions across aggregate TV genres. Calculated after model training, they explain each feature's influence on predic-tions by accounting for interactions, offering both global and local interpretability. They reflect the expected marginal contribution of a feature when added to a model in random order, akin to payoff in a cooperative game (Merrick & Taly

Table 3 Dialogue feature scores by TV series based on average robustness check between Elastic Net and Gradient Boosting

TV Series	Avg.R^2 Between Elastic Net and Gradient Boosting	1st feature	2nd feature	3rd feature	4th feature	5th feature	6th feature	7th feature	8th feature	9th feature	10th feature
2 Broke Girls	0.419	Feel_1 (0.67)	WPS_1 (− 2.07)*	Word Count_1 (1.20)	Engaged_1 (1.50)	Discrep_2 (1.38)	WPS_3 (0.37)	Dic_3 (0.81)	Clout_3 (0.94)	Six Letter Words_3 (− 3.35)***	Engaged_3 (1.79)*
30 Rock	0.455	Dic_1 (− 0.33)	Hear_1 (0.35)	WPS_1 (− 1.84)*	Dic_2 (1.07)*	Insight_2 (− 1.02)*	Fear_3 (− 0.55)	Certain_3 (− 1.13)*	WPS_3 (− 1.22)	WPS_1 (− 1.84)*	WPS_2 (− 3.72)**
A Different World	0.491	Drives_1 (− 3.85)	Reward_1 (− 3.70)	SD_scaled_1 (8.41)*	Joy_1 (4.34)	WPS_1 (7.97)**	Feel_2 (7.11)*	Anger_3 (− 11.022)***	Analytic_3 (0.71)	Six Letter Word_3 (12.49)**	Dic_3 (− 2.99)
A Million Little Things	0.433	Fear_1 (− 0.43)	Disgust_1 (− 0.53)	Power_1 (− 0.82)	WPS_1 (− 2.55)***	Percept_2 (− 0.96)	Hear_2 (− 0.34)	Analytic_2 (− 1.43)**	Feel_3 (− 1.21)	Anger_3 (0.03)	WPS_3 (− 0.54)
Adventure Time	0.654	Relativ_1 (− 0.21)	WPS_1 (− 2.01)*	Feel_1 (0.19)	Dic_1 (1.45)***	Space_1 (0.49)	WPS_2 (2.05)*	Dic_2 (− 0.39)	Cogproc_2 (− 0.71)*	Clout_3 (0.73)**	Six Letter Words_3 (− 0.98)**
Archer	0.546	Affiliation_1 (− 0.17)	Achieve_1 (− 0.10)	Drives_1 (− 0.40)	Affiliation_2 (− 0.50)	Engaged_2 (− 0.39)**	Disgust_2 (0.37)*	Clout_2 (− 0.53)**	Drives_2 (0.10)	Time_2 (− 0.47)**	Cogproc_3 (− 0.35)*
Arrow	0.487	Analytic_1 (0.99)***	Drives_1 (− 1.04)**	Word Count_1 (− 0.46)	Engaged_1 (− 0.40)	SD_scaled_2 (1.26)***	Word Count_2 (0.18)	Reward_3 (− 0.67)	Motion_3 (− 0.70)*	Word Count_3 (− 0.93)*	Drives_3 (− 0.45)
Becker	0.522	Cause_1 (− 1.53)	Neutral_1 (2.46)	Six Letter Word_1 (1.95)	Tentat_1 (1.26)	Word Count_1 (− 2.74)*	Dic_1 (− 2.99)	Joy_2 (1.58)	Disgust_2 (2.38)	Cause_3 (− 3.64)**	Netural_3 (2.72)*
Black−ish	0.631	Reward_1 (− 1.39)*	Hear_1 (2.21)**	Disgust_1 (1.58)*	WPS_1 (− 1.72)	Dic_1 (1.48)	WPS_2 (− 1.85)*	Dic_2 (− 2.33)**	Engaged_2 (− 1.43)*	Anger_3 (0.85)	Neutral (− 0.76)
Blindspot	0.762	Sadness_1 (− 1.05)	Joy_1 (− 0.16)	Achieve_1 (− 1.55)	Word Count_1 (− 1.31)	Clout_1 (1.93)**	Word Count_2 (− 0.86)	Sd_scaled_3 (1.30)	Fear_3 (0.81)	Achieve_3 (− 1.95)*	Word Count_3 (− 2.49)*
Breaking Bad	0.703	Word Count_2 (− 1.13)	Digust_2 (− 1.39)	Affiliation_1 (0.44)	SD_scaled_1 (− 2.24)	Feel_3 (2.79)**	Time_2 (0.43)	See_1 (1.74)	WPS_2 (2.20)	Cogproc_3 (− 2.13)	Neutral_2 (1.97)
Bull	0.771	Fear_1 (− 1.22)	WPS_1 (3.79)**	Dic_1 (0.44)	Time_2 (− 1.84)	Affiliation_2 (3.49)**	Insight_2 (− 0.43)	Engaged_2 (− 3.74)**	Dic_2 (0.27)	Cause_3 (3.52)**	Tentat_3 (− 3.15)**
Cold Case	0.594	SD_Scaled_1 (0.84)	Hear_2 (− 0.77)	Certain_2 (− 1.64)	Analytic_2 (− 1.41)	WPS_3 (2.66)*	Discrep_3 (− 1.64)	Neutral_3 (− 1.97)*	Fear_3 (2.20)*	Word Count_3 (− 5.54)***	Dic_3 (1.36)
Criminal Minds	0.463	Anger_1 (− 0.62)	Joy_1 (− 2.60)	Word Count_1 (− 2.49)	Engaged_1 (− 0.91)	Tone_2 (− 4.12)**	WPS_2 (− 6.28)***	See_3 (3.33)*	Percept_3 (− 2.88)	Joy_3 (− 5.19)***	Engaged_3 (− 3.59)*

Table 3 (continued)

TV Series	Avg.R^2 Between Elastic Net and Gradient Boosting	1st feature	2nd feature	3rd feature	4th feature	5th feature	6th feature	7th feature	8th feature	9th feature	10th feature
Crossing Jordan	0.597	Time_1 (− 1.12)	Sadness_1 (− 2.17)	WPS_1 (− 0.26)	Word Count_1 (3.01)*	Hear_2 (0.59)	WPS_2 (3.49)	Word Count_2 (1.35)	Anger_3 (− 1.37)	Authentic_3 (3.71)**	Clout_3 (− 2.09)
CSI: Crime Scene Investigation	0.665	Engaged_1 (2.24)	Word Count_1 (− 11.75)***	Fear_2 (− 3.89)	Neutral_2 (6.66)**	SD_scaled_2 (6.88)**	Word Count_2 (− 6.44)*	Analytic_2 (3.44)	Neutral_3 (− 0.71)	Analytic_3 (6.62)***	Word Count_3 (− 6.89)
CSI: NY	0.491	Neutral_1 (2.03)*	Joy_1 (− 2.71)**	Neutral_2 (2.69)*	Fear_2 (− 0.33)	Disgust_2 (3.55)**	Clout_2 (1.45)	Analytic_2 (0.53)	Positive_3 (− 0.59)	Joy_3 (− 1.62)	Neutral_3 (2.72)**
Deadwood	0.727	Time_1 (− 0.14)	Percept_1 (1.28)	Hear_2 (− 0.80)	Six Letter Words_2 (− 2.66)**	Relativ_3 (− 0.10)	Clout_3 (0.35)	Feel_3 (− 0.46)	WPS_3 (− 1.00)	Analytic_3 (− 0.18)	Six Letter Words_3 (− 50)
Dexter	0.472	Drives_1 (− 1.05)***	Clout_1 (0.30)	Disgust_2 (− 0.84)***	Authentic_2 (− 0.41)	WPS_2 (− 0.97)**	See_3 (0.64)*	Drives_3 (0.14)	Achieve_3 (− 0.31)	WPS_3 (− 1.24)***	Authentic_3 (− 0.29)
Empty Nest	0.749	Tentat_1 (− 9.41)**	Six Letter Words_1 (7.11)*	Dic_1 (− 8.93)***	Cogproc_1 (− 1.03)	Time_2 (9.22)***	WPS_2 (4.94)	Word Count_2 (6.77)*	Engaged_2 (5.15)	Word Count_3 (3.94)	WPS_3 (2.15)
Everybody Loves Raymond	0.531	Word Count_1 (− 0.58)	Neutral_1 (− 2.76)*	Dic_1 (11.29)***	Disgust_2 (− 2.02)	WPS_2 (5.37)*	Word Count_2 (− 4.51)*	WPS_3 (6.25)**	Hear_3 (3.73)**	Insight_3 (− 1.56)	Anger_3 (5.79)***
FBI	0.499	Space_1 (− 0.76)	Cause_1 (0.67)	Authentic_1 (− 0.06)	Reward_2 (− 1.31)*	Authentic_2 (− 0.50)	Authentic_3 (− 0.69)	SD_scaled_3 (0.16)	Cogproc_3 (− 0.06)	Engaged_3 (0.15)	Differ_3 (1.88)**
Grey's Anatomy	0.460	Relativ_1 (− 5.38)***	Tone_1 (− 5.84)***	Word Count_1 (− 8.16)***	Tone_2 (− 3.89)**	Six Letter Word _2 (3.32)	Word Count_2 (− 7.35)**	Space_3 (4.57)**	Disgust_3 (0.87)	Percept_3 (− 2.88)	Engaged_3 (− 7.57)***
Heroes	0.419	Fear_2 (− 4.40)	Insight_2 (− 0.31)	Fear_1 (− 1.08)	SD_Scaled_1 (2.75)	SD_Scaled_3 (1.53)	Authentic_1 (− 3.47)	Certain_2 (− 3.20)	Tone_3 (3.84)*	Clout_1 (− 1.29)	Cogproc_2 (− 0.13)
House	0.430	Motion_1 (5.80)*	Authentic_1 (− 2.18)	Disgust_1 (5.95)**	Authentic_2 (− 2.95)	Clout_2 (− 0.07)	Engaged_2 (3.16)	Insight_3 (− 0.58)	Certain_3 (− 5.35)**	Clout_3 (2.06)	Authentic_3 (− 2.57)
JAG	0.466	Neutral_2 (1.48)	Certain_2 (1.69)	Positive_2 (− 0.46)	Word Count_2 (− 1.12)	Drives_2 (− 0.17)	Word Count_3 (− 0.52)	Drives_3 (1.36)	Anger_3 (− 3.44)***	Clout_3 (3.12)**	Power_3 (1.10)
Las Vegas	0.529	WPS_1 (− 4.30)**	See_1 (2.14)	Word Count_1 (− 1.15)	Percept_2 (1.46)	Affiliation_2 (0.89)	Differ_2 (0.06)	Engaged_2 (5.06)***	Hear_3 (− 0.15)	SD_Scaled_3 (1.15)	Disgust_3 (− 1.82)
Law & Order	0.535	Anger_1 (− 4.31)**	See_1 (− 3.55)**	Disgust_1 (1.96)	Word Count_1 (9.25)***	Engaged_1 (− 8.11)***	Anger_2 (− 3.91)*	Disgust_2 (4.63)***	Neutral (2.12)	Risk_2 (− 5.45)**	WPS_3 (9.56)***

Table 3 (continued)

TV Series	Avg.R² Between Elastic Net and Gradient Boosting	1st feature	2nd feature	3rd feature	4th feature	5th feature	6th feature	7th feature	8th feature	9th feature	10th feature
Life	0.648	Affiliation_1 (0.54)	Reward_1 (1.37)	Affiliation_2 (2.66)*	Positive_2 (− 0.84)	Analytic_2 (− 0.33)	Space_2 (− 2.03)*	Drives_2 (− 0.25)	Power_3 (2.40)*	Drives_3 (− 0.49)	Achieve_3 (− 1.89)*
Life in Pieces	0.741	Relativ_1 (0.20)	Disgust_1 (− 1.11)	SD_Scaled_1 (− 0.20)	Power_1 (0.57)	Cogproc_1 (− 1.16)	Sadness_2 (1.85)*	Tone_2 (1.09)	Six Letter Word_3 (0.91)	Cause_3 (− 1.10)	Anger_3 (0.5)
Madam Secretary	0.871	Affiliation_1 (0.09)	Achieve_2 (− 1.63)	Positive_2 (− 1.04)	Risk_2 (3.19)**	WPS_2 (− 0.54)	Word Count_2 (− 2.99)	Risk_3 (− 1.23)	WPS_3 (− 2.17)	Word Count_3 (− 1.71)	Engaged_3 (− 3.26)
Medium	0.482	Cause_1 (− 2.28)*	SD_Scaled_1 (1.90)*	Word Count_1 (− 1.71)	Anger_1 (2.95)*	Motion_2 (− 1.69)*	Anger_2 (2.40)*	Dic_2 (− 1.96)*	Tone_3 (4.35)***	Clout_3 (− 2.00)*	Anger_3 (2.42)*
NCIS	0.620	Anger_1 (− 0.90)	Six Letter Words_1 (2.28)	Dic_1 (− 1.05)	Six Letter Words_2 (4.10)**	Engaged_2 (− 3.33)*	Analytic_2 (3.44)*	Anger_2 (− 3.04)*	Word Count_2 (− 4.40)**	Anger_3 (− 2.71)*	Word Count_3 (− 3.78)***
NCIS: Los Angeles	0.463	Analytic_1 (2.77)*	Word Count_1 (− 2.84)	Dic_1 (− 4.71)**	Engaged_1 (− 1.54)	Six Letter Word_2 (3.75)*	Word Count_2 (− 6.76)***	Dic_2 (− 2.75)	Engaged_2 (− 1.72)	Clout_3 (− 1.66)	Dic_3 (− 2.02)
NCIS: New Orleans	0.663	Anger_1 (− 3.68)**	Cause_1 (− 1.28)	Analytic_1 (2.04)	Dic_1 (− 1.20)	Percept_2 (3.68)**	Tone_2 (− 4.51)***	Analytic_2 (1.44)	Engaged_2 (− 4.92)***	Cogproc_2 (0.50)	Dic_2 (− 1.46)
NYPD Blue	0.623	Time_1 (− 5.30)***	Differ_1 (− 4.89)***	WPS_1 (− 8.06)***	Engaged_1 (3.72)*	Engaged_2 (0.38)	WPS_2 (2.92)	Word Count_2 (− 7.84)**	Time_2 (− 1.85)	WPS_3 (− 8.51)***	Word Count_3 (− 2.90)
Parks and Recreation	0.557	WPS_1 (0.42)	Word Count_1 (− 1.26)***	Six Letter Word_1 (− 0.63)	Percept_2 (− 0.46)	Certain_2 (0.53)	WPS_2 (− 0.32)	Analytic_2 (− 0.73)*	Word Count_2 (− 0.45)	Disgust_3 (0.88)*	Word Count_3 (− 1.23)***
Person of Interest	0.523	Achieve_1 (1.01)	Disgust_1 (− 1.89)	Risk_2 (3.06)*	Cause_2 (− 1.87)	Clout_2 (0.84)	Surprise_3 (3.10)*	Cause_3 (− 2.63)	Differ_3 (− 2.92)*	Six Letter Words_3 (− 4.85)**	Dic_3 (0.45)
Roseanne	0.609	Joy_1 (− 8.04)	Positive_1 (− 5.20)*	See_1 (− 3.28)	Anger_1 (6.05)	WPS_1 (− 13.27)**	Cause_1 (16.12)***	Word Count_1 (− 10.11)**	Analytic_2 (− 2.59)	Anger_2 (0.94)	Word Count_3 (− 5.89)*
Rules of Engagement	0.558	Feel_1 (− 1.15)	Insight_1 (− 1.83)	Joy_1 (1.09)	Fear_1 (− 3.46)***	Sadness_2 (− 1.42)	Positive_2 (0.87)	Affiliation_2 (− 1.64)	Motion_3 (1.34)	Tone_3 (− 1.28)	Dic_3 (3.03)**
Scorpion	0.723	Cogproc_1 (− 1.19)	WPS_1 (0.80)	Certain_1 (− 1.03)	Tentat_2 (2.71)	Disgust_2 (− 2.54)*	Affiliation_2 (− 1.33)	Cause_2 (− 0.94)	Dic_2 (3.91)**	Drives_3 (− 1.10)	Word Count_3 (− 3.64)**
Seinfeld	0.482	Six Letter Word_1 (− 7.30)*	Anger_1 (− 6.61)*	Positive_1 (2.55)	Relativ_1 (− 0.51)	Authentic (− 0.10)	Word Count_2 (− 0.95)	Dic_2 (− 9.75)**	Dic_3 (− 18.78)***	Sadness_3 (3.95)	Tone_3 (5.56)*

Table 3 (continued)

TV Series	Avg.R^2 Between Elastic Net and Gradient Boosting	1st feature	2nd feature	3rd feature	4th feature	5th feature	6th feature	7th feature	8th feature	9th feature	10th feature
Shark	0.634	Analytic_1 (− 1.01)	See_1 (− 0.44)	Engaged_1 (2.19)	WPS_1 (− 2.34)	Joy_2 (2.79)	WPS_2 (− 1.06)	Power_3 (− 0.37)	Affiliation_3 (− 0.38)	Anger_3 (− 0.69)	WPS_3 (1.68)
The Good Wife	0.703	Percept_1 (2.10)*	Sadness_1 (1.20)*	Feel_1 (0.37)	Sadness_2 (0.34)	Space_2 (0.63)	Analytic_2 (1.45)	Word Count_2 (− 2.02)**	Analytic_3 (1.04)	Word Count_3 (− 1.76)*	Positive_3 (− 1.82)**
The Nanny	0.421	Clout_1 (− 0.73)	Positive_1 (− 3.18)	Dic_1 (− 4.18)*	Affiliation_2 (− 1.67)	Cause_2 (− 4.93)**	Analytic_2 (3.29)	Dic_2 (− 0.98)	Clout_2 (− 0.76)	Drives_2 (− 4.23)	Dic_3 (− 2.42)
The Shield	0.440	Power_1 (1.63)*	Joy_1 (0.56)	Sadness_2 (0.77)	Authentic_1 (0.56)	Insight_1 (− 1.10)	Engaged_1 (0.06)	Word Count_1 (− 0.28)	Joy_2 (0.98)*	Sadness_3 (0.77)	Surprise_3 (1.09)
True Blood	0.465	Tone_1 (0.60)	Power_1 (1.05)	Anger_1 (0.99)	Risk_1 (0.04)	Drives_1 (− 0.29)	Six Letter Words_1 (1.79)**	Disgust_2 (− 1.82)**	SD_scaled_2 (0.49)	Anger_3 (0.90)	Disgust_3 (0.02)
Ugly Betty	0.700	Word Count_1 (− 1.21)	Tone_1 (− 2.86)*	Neutral_1 (4.37)**	Analytic_1 (0.86)	Cogproc_2 (0.32)	Analytic_2 (1.52)	Differ_2 (− 3.09)	Dic_2 (− 3.26)*	Tentat_3 (− 0.56)	Analytic_3 (2.90)
Without a Trace	0.520	WPS_1 (− 2.45)	Hear_1 (− 0.44)	Anger_1 (− 3.99)***	Disgust_2 (3.56)**	Discrep_2 (− 3.67)**	Word Count_2 (− 1.36)	Word Count_3 (− 0.65)	WPS_3 (− 8.65)*	Six Letter Words_3 (6.13)***	Word Count_3 (− 0.65)
Young Sheldon	0.907	Words Per Sentence_2 (3.26)*	Analytic_3 (1.44)	Discrep_3 (− 0.87)	Six Letter Words_3 (1.88)	Engaged_2 (− 2.10)*	Discrep_2 (− 2.07)**	Word Count_3 (− 2.74)**	Word Count_1 (− 2.39)*	Dic_3 (− 0.43)	Word Count_2 (− 0.99)

$^{*} = p < 0.05$

$^{**} = p < 0.01$

$^{***} = p < 0.001$

Table 4 Model evaluation scores and feature importances for individual TV genres

Genre	Comedy	Action & Adventure	Fantasy & Sci-Fi	Crime & Thriller	Horror & Mystery	Drama
Number of episodes	9,140	1,816	1,326	5,549	388	3,971
Elastic Net R^2	25.30%	32.23%	48.41%	46.88%	0.02%	28.74%
Gradient Boosting R^2	40.31%	43.37%	48.79%	66.76%	13.15%	41.74%
XG Boost R^2	40.09%	36.91%	46.21%	64.40%	13.57%	35.61%
XG Boost R^2 for Top Ten Features	44.43%	45.70%	44.93%	67%	66%	43.66%
1st SHAP value	Word Count_1 (.173) (1.67)	Anger_2 (0.03) (0.80)	Anger_1 (− 0.40) (1.19)	Engaged_1 (0.38) (0.92)	Joy_3 (− 0.01) (0.63)	Affiliation_1 (0.02) (0.65)
2nd SHAP value	Positive_2 (0.010) (.0.81)	Neutral_2 (− 0.09) (0.59)	Words Per Sentence_3 (− 0.10) (0.64)	Words Per Sentence_2 (0.16) (0.83)	Certain_1 (− 0.12) (0.30)	Word Count_3 (− 0.01) (0.64)
3rd SHAP value	Word Count_2 (− 0.01) (0.75)	Word Count _1 (0.03) (0.56)	Power_3 (− 0.11) (0.58)	Anger_2 (0.03) (0.75)	Engaged_2 (0.03) (0.27)	Word Count_2 (− 0.09) (0.56)
4th SHAP value	Disgust_1 (− 0.00) (0.58)	Six Letter Word_1 (− 0.00) (0.38)	Motion_1 (− 0.04) (0.43)	Anger_1 (− 0.01) (0.47)	Word Count _2 (− 0.09) (0.26)	Authentic_2 (− 0.10) (0.47)
5th SHAP value	Neutral_1 (0.07) (0.52)	Words Per Sentence_1 (0.00) (0.38)	Relativ_1 (0.01) (0.41)	Joy_1 (− 0.03) (0.38)	Discrep_1 (0.04) (0.22)	Fear_2 (− 0.15) (0.47)
6th SHAP value	Disgust_2 (0.02) (0.49)	Neutral_3 (− 0.05) (0.34)	Tone_1 (− 0.03) (0.27)	Word Count_2 (0.18) (0.37)	Cause_1 (0.01) (0.21)	Certain_2 (− 0.09) (0.44)
7th SHAP value	Words Per Sentence_1 (0.07) (0.47)	Six Letter Words_3 (0.03) (0.25)	Cognitive Process_1 (− 0.01) (0.24)	Anger_3 (0.00) (0.36)	Engaged_3 (0.02) (0.20)	Differ_1 (− 0.07) (0.43)
8th SHAP value	Drives_1 (0.07) (0.38)	Engaged_3 (0.02) (0.23)	Word Count_1 (− 0.02) (0.17)	Fear_3 (0.01) (0.25)	Surprise_3 (− 0.04) (0.15)	Certain_1 (0.04) (0.42)
9th SHAP value	Six Letter Words_1 (0.02) (0.37)	Achieve_1 (0.02) (0.21)	Neutral_1 (− 0.07) (0.16)	Neutral_2 (− 0.04) (0.20)	Motion_3 (− 0.04) (0.12)	Words Per Sentence_1 (− 0.10) (0.30)
10th SHAP value	Engaged_1 (0.05) (0.37)	Affiliation_2 (0.02) (0.20)	Risk (− 0.06) (0.16)	Reward_1 (0.07) (0.20)	Authentic_3 (− 0.02) (0.11)	Words Per Sentence_2 (− 0.06) (0.29)

2020). Shapley values improve transparency by clarifying why features matter.

The mean Shapley value captures a feature's average impact, including direction (positive or negative), while the mean absolute Shapley value reflects only the magnitude of influence, ignoring direction. Both metrics are presented in Table 4, with mean Shapley values listed first for clarity.

The findings from Table 4 illustrate particular genre boundaries and elements that inform each one. It appears that anger is particularly relevant in genres such as action and adventure (Anger_2 (0.03) (0.80)), fantasy and science fiction (Anger_1 (− 0.40) (1.19)), and crime and thriller (Anger_2 (0.03) (0.75)). As previously illustrated, words per sentence and word count are remarkably important and make notable marginal contributions across comedy (Word Count_1 (0.173) (1.67), Word Count_2 (− 0.01) (0.75), Words Per Sentence_1 (0.07) (0.47)) drama (Word Count_3 (− 0.01) (0.64), Word Count_2 (− 0.09) (0.56), Words Per Sentence_1 (− 0.10) (0.30), Words Per Sentence_2 (− 0.06) (0.29)), action and adventure (Word Count _1 (0.03) (0.56)), and crime (Words Per Sentence_2 (0.16) (0.83); Word Count_2 (0.18) (0.37)). Moreover, positive and neutral

tone modulation are also important in comedy (Positive_2 (0.010) (0.0.81); Neutral_1 (0.07) (0.52)) and action (Neutral_2 (− 0.09) (0.59), Neutral_3 (− 0.05) (0.34)).

Feature selection enhances model performance by identifying the most predictive variables, reducing noise, and improving interpretability while preventing overfitting. It helps pinpoint the key drivers of viewership and other outcomes (Table 5). The top feature importances at the TV series level include Clout_3 (0.141), Analytic_2 (0.052), Achieve_3 (0.040), Clout_1 (0.038), Fear_1 (0.024), Cause_2 (0.021). In contrast, at the TV genre level, the top features include Word Count_3 (0.203), Word Count_2 (0.071), Discrep_2 (0.064), Cogproc_2 (0.059), and Analytic_3 (0.054) (Table 6). These are useful dialogue features to help understand how a TV series may shape itself but also how people may perceive a particular genre. Together, this provides support for H3.

Table 5 Examination of top feature importances for TV series
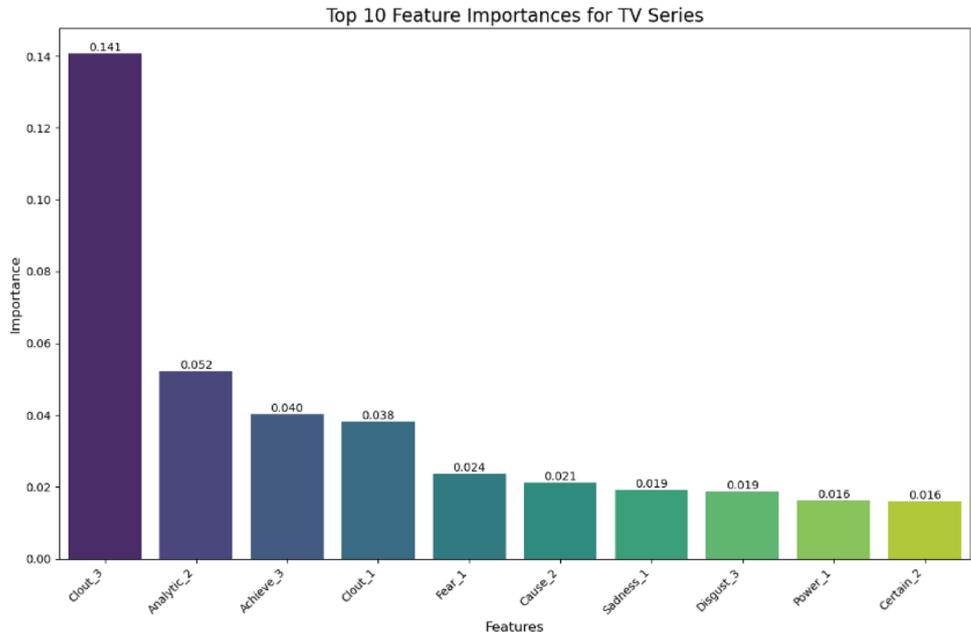
Table 6 Examination of top feature importances for TV genres
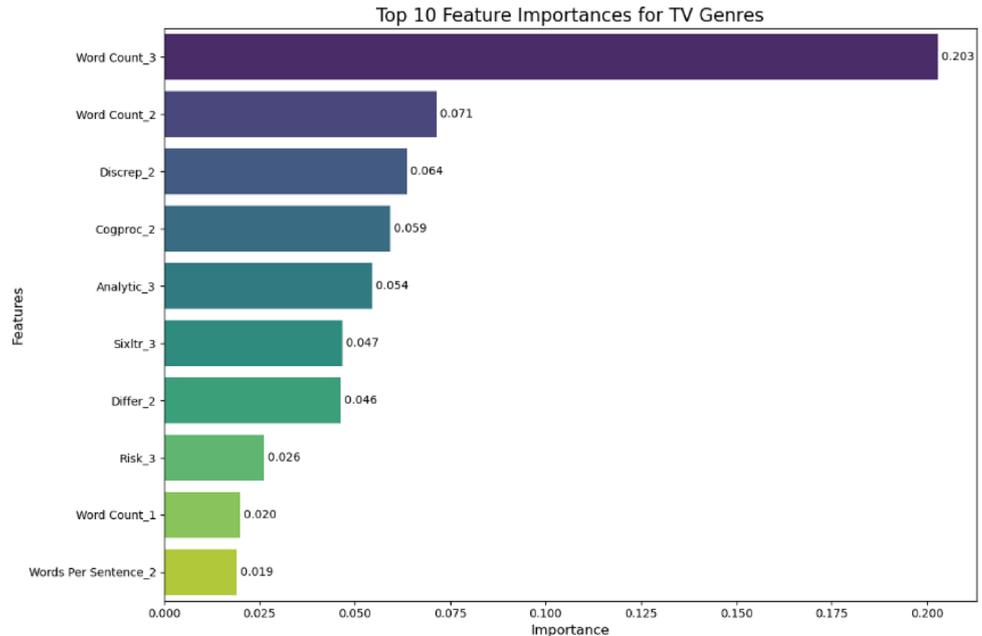
Discussion

This study extends peak-end theory (Fredrickson & Kahneman 1993) beyond its traditional domain of retrospective memory judgments into the realm of prospective behavioral prediction in serialized media. While past research emphasizes how emotionally salient moments shape recall and satisfaction, our results show that emotional peaks (e.g., Anger_2, Sadness_2) and resolution cues (e.g., Clout_3, Joy_3) embedded in dialogue can predict continued viewership, a future-oriented outcome. This reframes peak-end theory as not only applicable to how viewers evaluate experiences but also how they choose to re-engage. It introduces a longitudinal, behavioral dimension, suggesting that affective peaks and endings in narrative language act as predictive cues for audience retention. This challenges the notion that peak-end effects are limited to short-term, individual-level experiences, positioning them instead as structural narrative elements that shape collective audience behavior over time.

This study also advances seriality and engagement theory by shifting focus from structural plot devices (e.g., cliffhangers, arcs) to linguistic and cognitive mechanisms within dialogue. Traditional accounts emphasize narrative form or character development (Cochrane, 2014; Newman, 2024), but our findings reveal that linguistic complexity and cognitive activation, measured through features like words per sentence and cognitive-process terms (e.g., insight, cause), vary by genre and predict episode-to-episode viewership. This suggests audiences engage not only with what happens next but with how much mental effort an episode demands. Serial engagement thus stems from dialogic immersion, where cognitive labor becomes a psychological hook. In doing so, we propose a text-centered, computational model of seriality, integrating cognitive processing with long-form narrative retention, offering a fresh lens beyond traditional resolution-based models.

The emotional trajectory of TV dialogue, especially at peak moments and episode endings, clearly correlates with audience retention and future viewership. Serialized shows with strong emotional arcs (e.g., *Deadwood*, *Young Sheldon*, *Madam Secretary*) demonstrate greater predictive power, reinforcing the idea that compelling dialogue sustains engagement. In contrast, comedies (e.g., *30 Rock*, *Frasier*, *Modern Family*) and action-driven series (e.g., *24*, *Bones*, *The Walking Dead*) follow different, less predictable patterns. Moreover, different genres responded best to different modeling approaches: elastic net captured linear relationships, while gradient boosting identified nonlinear patterns and higher-order effects. These differences underscore the importance of model selection based on genre and analytical goals. Dialogue features also proved

central, as Words Per Sentence and Word Count help balance action with exposition, while Engagement, Cognitive Processing, Clout, and analytical language shape narrative intensity and viewer attention as characters reflect, problem-solve, and propel the plot.

At the aggregated TV genre level, Shapley values helped identify the most predictive features. XGBoost emerged as the top-performing model, consistent with prior findings on its effectiveness (Sheng & Yu, 2022; Udandarao & Gupta 2024). Genres such as fantasy, science fiction, crime, and thrillers showed high predictiveness, indicating they are strongly script-driven, with text features central to their success. In contrast, comedies may benefit from more verbose scripts, while action and adventure genres seemed to rely more on emotional intensity. Across models, Word Count remained a key predictor, though often ranking below emotional features like Fear, Reward, and Drive. This underscores the need to distinguish between features with strong predictive power and those offering marginal contributions when interpreting model outputs.

These findings reveal a clear distinction between feature importance at the TV series versus TV genre level. For individual series, features like narrative control (e.g., Clout), analytical thinking, and achievement-related language were most predictive of viewership, while at the genre level, Word Count, Words Per Sentence, and Cognitive Processing mattered more. This suggests that tonal shifts and linguistic complexity shape genre-level trends, whereas motivational and cognitive markers drive show-specific success. These differences highlight a dual challenge for showrunners and strategists: optimizing dialogue to boost individual series performance while aligning with genre norms. Platforms can use series-level features to enhance viewership and genre-level features to refine marketing and positioning. This framework offers a structured lens to understand how language shapes both show success and audience expectations.

Managerial implications

These findings have direct implications for the field of business analytics, particularly in applying predictive modeling to early-stage creative content. As data science capabilities expand within entertainment and media, there is growing demand for tools that transform unstructured narrative data into actionable insights. This study demonstrates that features embedded in dialogue, including emotional tone, linguistic structure, and narrative momentum, can be modeled to forecast audience retention. These predictive signals enable creative and strategy teams to assess content viability before release, helping to inform greenlighting, positioning, and scheduling decisions.

Beyond content evaluation, this approach supports benchmarking across series and genres, surfacing

storytelling patterns, and evolving audience preferences. Predictive features such as Conflict, Surprise, or Clout can also guide targeted marketing strategies, aligning brand identity with script tone, tailoring trailer or ad messaging, and timing campaigns around emotionally resonant episodes. For media firms and marketers alike, this framework offers a scalable, data-driven method for aligning creative output with consumer engagement.

While this study centers on predictive modeling, it also underscores practical considerations for scaling NLP and shaping creative workflows. Although model performance was strong, refining act segmentation and incorporating character-level analysis could improve accuracy. Future research might integrate critic reviews, social sentiment, and streaming metrics into a comprehensive performance dashboard. The findings also extend peak-end theory into marketing analytics: emotional intensity and dialogue structure at episode endings strongly predict future viewership. Marketers can spotlight these moments in trailers and ads to enhance memory and brand affinity, and time fan engagement around emotionally resonant scenes. Beyond prediction, these insights support upstream development—helping writers and showrunners refine arcs, pacing, and tone before production. Studios and platforms might use dashboards to guide writers' room discussions or prototype data-informed storytelling formats (Gupta 2001; Porter 2001).

Limitations and future research

While prior research (e.g., Reagan et al. 2016; Del Vecchio et al. 2021) has shown that NLP can identify narrative structures and emotional arcs, these studies focus more on structural patterns than on audience response. Detecting emotional arcs via NLP does not, by itself, indicate storytelling success from the viewer's perspective. This study uses LIWC, a validated tool for extracting emotional and psychological content, but its bag-of-words design overlooks semantic and syntactic relationships. Future work could adopt transformer-based models (e.g., BERT, GPT) to better capture narrative flow and character interactions.

Further research should integrate reception-based data, such as satisfaction scores, retention metrics, or qualitative feedback, to test whether identified patterns align with audience preferences. This study also did not compare NLP-derived features to human-coded benchmarks, given its exploratory aim of testing automated content analytics. Future work should incorporate such comparisons to evaluate accuracy and interpretability, enhancing the validity of NLP approaches in content research.

Appendix 1: NLP Dictionary

To access this dashboard, it is essential to understand how features and variables are coded. The following definitions outline the key terms and measurements used in the analysis.

Story act designations

Designations _1, _2, _3 correspond to act one, act two, and act three within a given TV episode. These acts were determined based on script length to examine how emotions varied across the story structure.

Emotion and psychological engagement metrics

- **Sd_Scaled**: The standard deviation is scaled across all emotions for comparative analysis, derived from the sum of standard deviations per act. This approach provides insight into emotional variance across different story segments. It is based on a transportation measurement that assesses the speed and depth of a viewer's immersion in a story—both mentally and emotionally. This metric captures how quickly audiences lose awareness of their surroundings and become absorbed in the narrative. By dividing the standard deviation of overall emotion scores by the mean, it quantifies shifts, variability, and intensity of emotions throughout the experience.
- **Anger**: The level of anger present in an act.
- **Surprise**: The level of surprise present in an act.
- **Disgust**: The level of disgust present in an act.
- **Sadness**: The level of sadness present in an act.
- **Neutral**: The level of neutrality present in an act.
- **Fear**: The level of fear present in an act.
- **Joy**: The level of joy present in an act.
- **Positive**: The amount of positive sentiment present in an act.
- **Engaged**: High psychological involvement or emotional investment, characterized by increased use of personal pronouns, emotional words, and cognitive processing.

Linguistic and narrative complexity metrics

- **Word Count**: Total number of words in an act.
- **Analytic**: The presence of analytical, formal, or logical discussion.

- **Clout**: The extent of social status, confidence, or leadership discourse.
- **Authenticity**: The presence of honest, non-filtered, or unregulated discussion.
- **Tone**: A measure of positivity; scores below 50 indicate negativity.
- **WPS (Words per Sentence)**: The average number of words per sentence in an act.
- **Six Letter Word**: Percentage of words with more than six letters.
- **Dic (Dictionary)**: Percentage of words captured in the dictionary.

Cognitive processing metrics

- **Cogprocess**: Aggregate measure of words indicating active information processing and mental activity, including causation.
- **Insight**: Words that indicate realizations or understanding.
- **Cause**: Words that signal causal relationships between elements.
- **Discrep**: Words that indicate counterfactual thinking (e.g., should, could, would).
- **Tentative**: Words that indicate uncertainty or possibility (e.g., maybe, perhaps).
- **Certain**: Words that indicate absolute statements (e.g., always, never).
- **Differ**: Words that indicate differentiation between concepts (e.g., but, else).

Perceptual processing metrics

- **Perceptual**: Aggregate measure of words describing perception (e.g., look, heard, feel).
- **See**: Words associated with visual perception.
- **Hear**: Words associated with auditory perception.
- **Feel**: Words related to tactile sensation.

Motivational drives metrics

- **Drives**: Aggregate measure of different motivations expressed in an act.
- **Affiliation**: Words associated with social relationships (e.g., ally, friend).
- **Achieve**: Words related to success and accomplishment.
- **Power**: Words related to dominance and hierarchical structures.
- **Reward**: Words related to receiving benefits or prizes.
- **Risk**: Words related to danger or uncertainty.

Spatial and temporal metrics

- **Relativ (Relativity)**: Aggregate measure of words describing spatial relationships (e.g., area, bend, exit).
- **Motion**: Words related to movement (e.g., arrive, go, car).
- **Space**: Words indicating spatial direction (e.g., down, in).
- **Time**: Words related to time duration (e.g., end, until, season).

Declarations

References

Akiba, T., S. Sano, T. Yanase, T. Ohta, & M. Koyama. 2019. Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International conference on knowledge discovery and data mining* (*KDD '19*), 2623–2631.

Allen, D. M. 1974. The relationship between variable selection and data augmentation and a method of prediction. *Technometrics* 16 (2): 125–127.

Amit, Y., and D. Geman. 1997. Shape quantization and recognition with randomized trees. *Neural Computation* 9:1545–1588.

Balducci, B., and D. Marinova. 2018. Unstructured data in marketing. *Journal of the Academy of Marketing Science* 46:557–590.

Baumgartner, H., M. Sujan, and D. Padgett. 1997. Patterns of affective reactions to advertisements: The integration of moment-to-moment responses into overall judgments. *Journal of Marketing Research* 34 (2): 219–232.

Berger, J., and K. L. Milkman. 2012. What makes online content viral? *Journal of Marketing Research* 49 (2): 192–205.

Bietti, L., O. Tilston, and A. Bangerter. 2018. Storytelling as adaptive collective sensemaking. *Topics in Cognitive Science* 11 (4): 710–732.

Bourdon, J. 2021. *Measuring TV audiences: from people meters to big data*. Media, Culture & Society.

Boyd, R. L., K. G. Blackburn, and J. W. Pennebaker. 2020. The narrative arc: Revealing core narrative structures through text analysis. *Science Advances* 6 (32): 1–9.

Breiman, L. 1996. Bagging predictors. *Machine Learning* 24 (2): 123–140.

Cochrane, T. 2014. Narrative and character formation. *The Journal of Aesthetics and Art Criticism,* 72 (3): 303–315.

Cohen, J. 2001. Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Communication and Society* 4 (3): 245–264.

Dale, K., J. Fisher, J. Liao, and E. Grinberg. 2023. The shape of inspiration: Exploring the emotional arcs and self-transcendent elicitors within inspirational movies. *Media Psychology* 26 (6): 767–789.

Del Vecchio, M., A. Kharlamov, G. Parry, and G. Pogrebna. 2021. Improving productivity in Hollywood with data science: Using emotional arcs of movies to drive product and service innovation in entertainment industries. *Journal of the Operational Research Society* 72 (5): 1110–1137.

Dewi, C., and R. Chen. 2019. Random forest and support vector machine on features selection for regression analysis. *International Journal of Innovative Computing, Information and Control* 15 (6): 2027–2037.

Dharaniya, R., J. Indumathi, and V. Kaliraj. 2023. A design of movie script generation based on natural language processing by optimized ensemble deep learning with heuristic algorithm. *Data & Knowledge Engineering* 146 : 102150.

Eliashberg, J., S. Hui, and Z. Zhang. 2014. Assessing box office performance using movie scripts: A kernel-based approach. *IEEE Transactions on Knowledge and Data Engineering* 26 (11): 2639–2648.

Fredrickson, B. L., and D. Kahneman. 1993. Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology* 65 (1): 45–55.

Geiger, S., and J. Newhagen. 1993. Revealing the black box: Information processing and media effects. *Journal of Communication* 43 (4): 42–50.

Geisser, S. 1975. The predictive sample reuse method with applications. *Journal of the American Statistical Association* 70 (350): 320–328.

Géron, A. 2019. *Hands-on machine learning with Scikit-Learn, Keras & TensorFlow*, 2nd ed. O'Reilly.

Graesser, A. C., D. S. McNamara, Z. Cai, M. Conley, H. Li, and J. Pennebaker. 2014. Coh-metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal* 115 (2): 210–229.

Green, M. C., and T. C. Brock. 2000. The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology* 79 (5): 701–721.

Green, M. C., and J. L. Clark. 2013. Transportation into narrative worlds: Implications for entertainment media influences on tobacco use. *Addiction* 108 (3): 477–484.

Gross, J., and K. Desveaud. 2024. Navigating computational linguistic in marketing practices: The barriers of natural language processing in social media marketing and a path to future research. *Journal of Marketing Analytics*. Advance online publication.

Gunasegaran, T., and Y. N. Cheah. 2017. Evolutionary cross validation. *In 2017 8th International Conference on Information Technology (ICIT)* (pp. 89–95). IEEE.

Gupta, A. 2001. Business unit strategy: The quest for competitive advantage. In *The portable MBA in strategy*, 2nd ed., ed. L. Fahey and R. Randall, 62–86. Wiley.

Hair, J., W. Black, B. Babin, and R. Anderson. 2018. *Multivariate data analysis*, 8th ed. Cengage Learning EMEA.

Hall, A. E. 2003. Reading realism: Audiences' evaluations of the reality of media texts. *Journal of Communication* 53:624–641.

Horowitz, A. G., K. McCarthy, and S. Sen. 2024. A review of the peakend rule in mental health contexts. *Current Opinion in Psychology* 58: 101845.

Kahneman, D., B. L. Fredrickson, C. A., Schreiber, and D. A. Redelmeier. 1993. When More Pain Is Preferred to Less: Adding a Better End. *Psychological Science*, 4 (6): 401–405. http://www.jstor.org/stable/40062570.

Kim, S. 2018. Audience measurement and analysis. In *Handbook of Media Management and Economics*, ed. A. Albarran, B. Mierzejewska, and J. Jung, 379–393. Routledge.

Kursa, M. B., and W. R. Rudnicki. 2011. The all relevant feature selection using random forest. *arXiv preprint* arXiv:1106.5112.

Lu, A. S., M. C. Green, C. V. Sousa, J. Hwang, I. M. Lee, D. Thompson, & T. Baranowski. 2025. To pause with a cliffhanger or a temporary closure? The differential impact of serial versus episodic narratives on children's physical activity behaviors. *Communication Research* 52 (5): 600–625.

Leith, A. P. 2021. Parasocial cues: The ubiquity of parasocial relationships on Twitch. *Communication Monographs* 88 (1): 111–129.

Luchen, F., and L. Zhongwei. 2023. ChatGPT begins: A reflection on the involvement of AI in the creation of film and television scripts. *Frontiers in Art Research* 5 (17): 1–6.

Martinez, V. R., K. Somandepalli, K. Singla, A. Ramakrishna, Y. T. Uhls, and S. Narayanan. 2019. Violence rating prediction from movie scripts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 671–678.

McKee, R. 1997. *Story: Substance, structure, style and the principles of screenwriting*. HarperCollins.

Merrick, L., and A. Taly. 2020. The explanation game: Explaining machine learning models using shapley values. In *International cross-domain conference for machine learning and knowledge extraction*, 17–38. Springer.

Mittell, J. 2004. *Genre and television: From cop shows to cartoons in American culture*. Routledge.

Murthy, D., S. Bowman, A. J. Gross, and M. McGarry. 2015. Do we tweet differently from our mobile devices? A study of language differences on mobile and web-based Twitter platforms. *Journal of Communication* 65 (5): 816–837.

Nabi, R. L., C. R. Stitt, J. Halford, and K. Finnerty. 2006. Emotional and cognitive predictors of the enjoyment of reality-based and fictional television programming: An elaboration of the uses and gratifications perspective. *Media Psychology* 8 (4): 421–447.

Napoli, P. M. 2011. *Audience evolution: New technologies and the transformation of media audiences*. Columbia University Press.

Nelson, J. L., and J. G. Webster. 2016. Audience currencies in the age of big data. *International Journal on Media Management* 18 (1): 9–24.

Newman, S. J. 2024. Four Temporalities: Toward a Typology of Narrative Forms. *Sociological Theory,* 42 (4): 283–306.

Oliver, M. B. 2003. Mood management and selective exposure. In *Communication and emotion: Essays in honor of Dolf Zillmann*, ed. J. Bryant, D. Roskos-Ewoldsen, and J. Cantor, 85–106. Lawrence Erlbaum Associates.

Petrescu, M., and A. S. Krishen. 2023. A decade of marketing analytics and more to come: JMA insights. *Journal of Marketing Analytics* 11:117–129.

Porter, M. 2001. Competing across locations: Enhancing competitive advantage through a global strategy. In *The portable MBA in strategy*, 2nd ed., ed. L. Fahey and R. Randall, 87–108. Wiley.

Raj, S., A. Sharma, S. Saha, B. Singh, and N. Pedanekar. 2024. Transformative movie discovery. Large language models for recommendation and genre prediction. *IEEE Access* 12:186626–186638.

Ramakrishna, A., V. R. Martínez, N. Malandrakis, K. Singla, and S. Narayanan. 2017. Linguistic analysis of differences in portrayal of movie characters. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1669–1678). Association for Computational Linguistics.

Reagan, A., L. Mitchell, D. Kiley, C. Danforth, and P. Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science* 5 (31): 1–12.

Redelmeier, D. A., and D. Kahneman. 1996. Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain* 66 (1): 3–8.

Rubin, A. M. 2009. Uses-and-gratifications perspective on media effects. In *Media effects advances in theory and research*, ed. Bryant and Zillmann. Routledge.

Rucker, D. D., and R. E. Petty. 2004. Emotion specificity and consumer behavior: Anger, sadness, and preference for activity. *Motivation and Emotion* 28 (1): 3–21.

Scheiber, N., and J. Koblin. 2023. Will a chatbot write the next 'Succession'? *The New York Times*. https://www.nytimes.com/2023/04/29/business/media/writers-guild-hollywood-ai-chatgpt.html

Schibler, K., L. Hahn, and M. C. Green. 2024. Investigating responses to narrative cliffhangers using affective disposition theory. *Media Psychology* 27 (1): 1–25.

Selnow, G. W. 1986. Solving problems on prime-time television. *Journal of Communication* 36 (2): 63–72.

Shehu, E., T. H. A. Bijmolt, and M. Clement. 2016. The effects of likeability dynamics on consumers' intention to share online video advertisements. *Journal of Interactive Marketing* 35:27–43.

Sheng, C., and H. Yu 2022. An optimized prediction algorithm based on XGBoost. In *2022 international conference on networking and network applications (NaNA)* (pp. 1–6). IEEE.

Smith, C. A., and P. C. Ellsworth. 1985. Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology* 48 (4): 813–838.

Srivastava, S., S. Oberoi, and V. K. Gupta. 2023. The story and the storyteller: Strategic storytelling that captures human attention for entrepreneurs. *Business Horizons* 66 (4): 347–358.

Open Subtitles. 2025. *Home page*. Open Subtitles. Retrieved February 21, 2025, from https://www.opensubtitles.com/en/tvshows.

Tamborini, R., N. D. Bowman, A. Eden, M. Grizzard, and A. Organ. 2010. Defining media enjoyment as the satisfaction of intrinsic needs. *Journal of Communication* 60 (4): 758–777.

Tausczik, Y. R., and J. W. Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29 (1): 24–54.

Udandarao, V., & P. Gupta. 2024. Movie revenue prediction using machine learning models. *arXiv preprint* arXiv:2405.11651.

van Laer, T., K. de Ruyter, L. M. Visconti, and M. Wetzels. 2014. The extended transportation-imagery model: A meta-analysis of narrative persuasion. *Journal of Consumer Research* 40 (5): 797–817.

Wirz, D. S., A. Ort, B. Rasch, and A. Fahr. 2023. The role of cliffhangers in serial entertainment: An experiment on cliffhangers' effects on enjoyment, arousal, and intention to continue watching. *Psychology of Popular Media* 12 (2): 186.

Xu, Z., C. Vail, A. S. Kohli, and S. Tajdini. 2021. Understanding changes in a brand's core positioning and customer engagement: A sentiment analysis of a brand-owned Facebook site. *Journal of Marketing Analytics* 9 (1): 3–16.

Yates, D., and M. Z. Islam. 2021. Fastforest: Increasing random forest processing speed while maintaining accuracy. *Information Sciences* 557:130–152.

Yuan, H., and A. Hernandez. 2023. User cold start problem in recommendation systems: A systematic review. *IEEE Access* 11:136958–136977.

Anthony Palomba is a media scholar specializing in machine learning, causal inference, and applied econometrics. His research explores how audiences engage with media content and how technology reshapes competition in the media, advertising, and gaming industries. His work has been published in *Telematics and Informatics, Journal of Retailing and Consumer Services, Computers in Human Behavior, Entertainment Computing, Games & Culture*, and *International Journal on Media Management*. A former research manager at Ipsos, he has led projects for HBO, Facebook, Fox Sports 1, CNN, Sesame Street, and NBCUniversal. He previously conducted research with Nielsen during his doctoral studies.