

Instagram Reels Engagement Drivers: Machine Learning Insights into Celebrities and Brands

Elif Hasret KUMCU¹, & Vesile ÖZÇİFÇİ²

Abstract

The paper explores the major motivators of user interaction on Instagram Reels, a rising short video form in digital marketing. Despite the central role played by Reels in content strategies, there is still minimal empirical research on which specific features are relevant in driving the engagement. To fill this gap, we analyzed 438 Instagram Reels of global celebrities and brands in the period from January to March 2024. A combination of exploratory data analysis with a Random Forest regression model was applied in a specific way so that it could estimate the importance of the features and find the most significant predictors of the engagement rate. These were the ranked features, which then were the input of a K-Means clustering analysis which indicated patterns of content that were linked to different degrees of audience response. The results indicate that celebrity and brand presence is the most influential factor (22.10% importance), followed by content type (19.03%), age representation (16.27%), and human activity (16.16%). In contrast to common assumptions, the findings show that neutral emotional tone generates the highest engagement, while positive tone yields the lowest, suggesting that authenticity and subtle emotional expression may be more effective than overt positivity in short-form video contexts. The combination of machine learning methods and traditional engagement theory within the study gives practical recommendations to the marketer and will create a clear, reproducible map of research on short-form video analytics in the future.

Keywords: Instagram Reels, Digital marketing, Engagement Rate, Machine Learning, Social Media Marketing

Instagram Reels Etkileřimini Artıran Faktörler: Ünlüler ve Markalar Hakkında Makine Öğrenmesi İğröřüleri

Öz

Bu alışma, dijital pazarlamada yükselen kısa video formatlarından biri olan Instagram Reels üzerinde kullanıcı etkileřimini yönlendiren temel faktörleri incelemektedir. Reels içerik stratejilerinde merkezi bir rol oynamasına rağmen, etkileřimi hangi spesifik özelliklerin yönlendirdiğine dair ampirik arařtırmalar hâlen sınırlıdır. Bu boşluğu doldurmak amacıyla, Ocak–Mart 2024 döneminde küresel ölçekte tanınan ünlüler ve markalar tarafından paylaşılan 438 Instagram Reels içeriğİ analiz edilmiştir. Özelliklerin önemini tahmin edebilmek ve etkileřim oranının en güçlü belirleyicilerini ortaya koyabilmek için keřşisel veri analizi ile Random Forest regresyon modeli birlikte kullanılmıştır. Elde edilen özellik kümeleme analizine girdi olarak kullanılmıştır. Sonuçlar, ünlü ve marka varlığının en etkili faktör olduğunu (%22,10 önem düzeyi), bunu içerik türü (%19,03), yař temsili (%16,27) ve insan aktivitesi (%16,16) deėişkenlerinin izlediğini göstermektedir. Yaygın varsayımların aksine, bulgular nötr duygusal tonun en yüksek etkileřimi sağladığını, olumlu tonun ise en düşük etkileřim düzeyine sahip olduğunu ortaya koymaktadır. Bu durum, kısa video bağlamında özgünlük ve daha örtük duygusal ifadenin, açık ve yoğun pozitiflikten daha etkili olabileceğini düşündürmektedir. Makine öğrenmesi yöntemlerinin geleneksel etkileřim teorileriyle birlikte kullanılması, pazarlamacılara yönelik pratik öneriler sunmakta ve gelecekte kısa video analitiğİ alanındaki arařtırmalar için açık ve tekrarlanabilir bir çerçeve ortaya koymaktadır.

Anahtar Kelimeler: Instagram Reels, Dijital Pazarlama, Etkileřim Oranı, Makine Öğrenmesi, Sosyal Medya Pazarlaması

Please Cite As / Atıf İin:

Kumcu, E. H. & Özçifçi, V. (2026). Instagram reels engagement drivers: Machine learning insights into celebrities and brands. *Manas Sosyal Arařtırmalar Dergisi*, 15 (3), 1075-1094. <https://doi.org/10.33206/mjss.1879442>

Received Date / Geliř Tarihi: 02.02.2026

Accepted Date / Kabul Tarihi: 18.06.2026

¹ Asst. Prof. - Aksaray University, School of Technical Sciences Vocational College, elifhasretkumcu@aksaray.edu.tr,

 ORCID: 0000-0003-2732-7006

² Assoc. Prof. Dr. - Aksaray University, Faculty of Economics and Administrative Sciences, vesileozcifci@aksaray.edu.tr,

 ORCID: 0000-0002-8011-9137



Introduction

Instagram has become a prominent platform in digital marketing and user interaction, especially due to the increasing popularity of its short-content video, Reels. Instagram currently serves over 2 billion of its user worldwide, becoming a key platform where creators, brands, and other figures can find audiences at large scale (Statista, 2024). Reels have also proven to be significantly more interactive than the usual posts or stories on Instagram, which is explained by the fact that they are prioritized by the algorithm, occupy the full visual interface, and have the fastest cycle of content consumption. All these obvious benefits notwithstanding, the scientific insight of what exactly makes people engage in Reels is rather scarce. In previous research, user behavior on the social media is analyzed in fragmented forms such as individual content themes, motivations, or usage patterns (Meng & Leung, 2021:11; Omar & Dequan, 2020:130). Although these studies offer a starting point, they fail to sufficiently explain the multi-dimensional and intricate characteristics of engagement in the modern context of the short-form video. New media scholars have suggested alternative theories that can be used to describe a wider behavioral dynamic. As an illustration, based on the uses and gratifications theory, Shao (2009:4) theorizes user activity along with three axes; production (content creation), consumption (viewing), and participation (interaction). This multidimensional view directly applies to Reels where people not only consume the content but constantly communicate in the form of likes, shares, comments, and saves. Nevertheless, there is limited empirical evidence on the application of such frameworks to Instagram Reels.

Existing literature has largely focused on platforms such as Facebook, YouTube, and earlier Instagram formats, emphasizing factors such as content type, visual attractiveness, and message framing (Alhabash & Ma, 2017:7; Voorveld et al., 2018:50). Nevertheless, Reels on Instagram have significant differences in terms of design and experience-related features: vertical orientation of video, an extensive use of popular audio, accelerated storytelling, and the search of content suggested by the algorithm (Johnson & Ranzini, 2018:156). Such features of platforms are likely to influence engagement differently than they would be reflected in the results of older or structurally distinct platforms. Therefore, there is a clear gap in understanding how multiple content attributes jointly influence user engagement in short-form, algorithm-driven video environments such as Instagram Reels. In order to fill this gap, the current research adopts a machine learning-based methodology that combines the application of the Random Forest regression and K-Means clustering to investigate the determinants of engagement on Instagram Reels. To address this gap, the study adopts a machine learning-based approach to examine the determinants of engagement on Instagram Reels.

The study contributes to the theoretical discussion because it bases social media interaction on computational analysis, making the results of machine learning consistent with behavioral and communication models, including parasocial interaction, emotional contagion, and uses and gratifications. In doing so, it addresses the lack of integrative, data-driven approaches in prior literature and provides a more comprehensive understanding of engagement mechanisms in short-form video ecosystems. In practice, the results allow marketers and content developers to gain empirically supported knowledge on what video qualities best contribute to user reaction. In addition, the study provides a blueprint of effective methods of conducting research in the future because it created a replicable analytical framework to analyze the engagement with short-form video on platforms like TikTok, YouTube Shorts and Snapchat Spotlight. Based on the identified gap, the study addresses the following research questions: Which content attributes most strongly influence user engagement in Instagram Reels? How do different combinations of content features form distinct engagement patterns? To what extent do identity-related factors (e.g. celebrity presence) and emotional tone shape engagement outcomes in short-form video content?

Literature Review

The use of social media has taken a multidimensional nature and is still being researched. There is previous conceptual work demonstrating that the engagement is polysemic, and behavioural, cognitive, and emotional layers denote the various types of online interactions between users (Trunfio & Rossi, 2021). Their systematic review demonstrates that researchers have quantified engagement by a range of quantitative measures, normalised indexes, and qualitative measures often on the basis of the COBRA model and categorise behaviours into consumption, participation and creation. The recent studies build on these results to investigate the influence of brand-controlled communication on engagement. Meta-analysis indicates that branded social media content has an influence on sales performance, in addition to engagement, which disproves the notion that owned media is only going to enhance engagement (Liadeli, Sotgiu & Verlegh,

2023). The emotional content makes it more engaging and functional content makes it sell especially among brands with smaller audiences or in high-power distance markets. Another factor is engagement based on environmental and platform design. According to (Reimer, 2023), media type, brand involvement and the cross channel tactic affect the engagement in three connected environments; imposed, selected and created environment. The analysis of over 7,000 brand posts on Facebook, Instagram, and Twitter reveals that content structure, Twitter and Instagram platform dynamics, and brand activity have a significant influence on the degree of engagement.

Other than brand centric perspectives, psychological and motivational engagement aspects have also been researched by scientists. The relationship between recreational use of digital media and the psychological well-being of students is mediated by social media, and both cognitive and behavioural engagement influence (positively) and affective engagement (negatively) (Asan and Zwiendelaar, 2025). This implies that the well-being of the users is different depending on the type of engagement. Brand perceptions are also affected by customer generated content. Authentic communication, positive customer experience, and user-generated content have been identified by retailers as being effective to improve brand sentiment and loyalty when the brands have an open and meaningful interaction with the customers (Krisprimandoyo, Sufa, Wardani & Widiyanto, 2024). In luxury brand research, the complexity of engagement motivations is noted. The mixed methods are applied (Aldhamiri et al., 2024) to identify intrinsic (entertainment, aesthetics, authenticity, nostalgia) and extrinsic (social support, status seeking) determinants of active and passive participation. These intentions interact with the value orientation and brand recognition of customers to cause luxury content interaction. It is found in the literature that social media engagement occurs via conceptual frameworks, content strategies, platform structures, influencer traits, and user motives.

The fast development of social media platforms has changed the way marketing is done and patterns of consumer interaction. One product is Instagram Reels, a new feature designed to compete with TikTok in the short-form video format, which becomes one of the most popular platforms that content creators and brands can use to communicate with customers. With more than 2 billion active users worldwide, as of 2024, Instagram presents an enormous opportunity to marketers to manipulate consumer behavior (Statista, 2024). The implications of the recent findings are useful when it comes to public relations professionals and media strategists as Instagram Reels have become an important tool in modern-day marketing and communication campaigns.

As an example, Menon (2022) also found entertainment, novelty, and trendiness to be the main gratifications users of Reels are pursuing. It implies that fresh, entertaining, and trendy content should be among the priorities of media marketers and influencers who want to maximize their involvement. In spite of this increase, there is still a lack of scholarly research on Instagram Reels, in comparison with such platforms as Facebook, YouTube, or even TikTok. Even before the current study, previous studies have concentrated predominantly on image-based posts or long-form video types of engagement factors but have overlooked other types of short-form video (Alhabash and Ma, 2017:5; Voorveld et al., 2018:38). Nonetheless, Instagram Reels have their own exclusive characteristics, such as temporary viewing, delivery of content by an algorithm, using popular audio elements, and the presence of emotionally-charged and fast-paced content, which require a different analytical prism (Johnson & Ranzini, 2018:148).

These researches point out interactivity, content type, and aesthetic as the main drivers of engagement. The research is based on and deviates the available literature by paying attention to Reels. It broadens the theoretical and empirical knowledge of the contents of short-form videos that are interesting. Other studies like the one by Berger & Milkman (2012:8) highlight the emotive nature of material, especially how positive, awe-inspiring or surprising material is more prone to sharing. On the same note, Kietzmann et al. (2011:250) emphasise the significance of visual storytelling and the creation of communities in the digital space.

These theoretical viewpoints are complemented with the recent research about the content strategies on online platforms. As a case in point, Moran et al. (2020) and Lou & Xie (2021) demonstrate that engagement is caused by message features and entertainment value. Equally, Qin (2020), Wibowo et al. (2020) and Onofrei et al. (2022) study user interaction behavior and content attribute, and the mediating roles of the content quality and source credibility, respectively.

The influencers as the drivers of engagement are being studied in more studies. With an increase in the number of influencer followers, engagement will increase but eventually fall because of weaker relation ties (Wies, Bleier and Edeling, 2023). Authenticity and consumer engagement to influencer campaigns are enhanced with content customisation and low brand familiarity.

According to the latest research, there is massive growth in applying ML to enhance digital and social media marketing. Current studies indicate that the conventional engagement metrics are unable to capture more in-depth behavioural signals, whereas AI-driven analytics (especially those that apply both ML and NLP) can be more helpful in terms of understanding sentiment, pattern, and consumer responses in large volumes of data (Reddy et al., 2021). The sentiment and topic models of ML are superior to traditional forecasting trend and real-time strategic decision-making. ML is being applied more often to predict non-data customer engagement. Schivinski (2021) demonstrates the way in which a conditional inference tree architecture can be used to identify the cognitive processes that affect social media consumption, participation, and invention with the help of the ML. There are 5 interaction guidelines and 5 consumer subtypes that offer data-based segmentation methods to the marketer.

Twitter interaction is predicted by machine learning classifiers, such as multinomial logistic regression, decision trees, k-NN and random forest. The interactive aspects, defined by vividness cues, and time variables forecasted e-commerce posts in a big dataset, and the random forest model was more effective than other singular classifiers (Andariesta & Wasesa, 2023). ML enhances brand strength and consumer participation when used as a social media strategy. The adoption of ML, personalisation, engagement and brand strength are all positively related in UAE retail, but brand variety of the platform moderates this relationship (Al Kurdi et al., 2024). ML enhances the effectiveness of marketing by personalisation.

ML is scalable in the content-driven applications of visual feature analysis. Latest frame-by-frame video coding indicates that social media films involving human faces increase the interaction, which is limited by face length and brand placement (Wallach et al., 2025). ML is able to read tiny visual prompts that affect user behaviour in bulk. These works show that ML alters digital marketing, enhancing the analytics, predicting behaviour, enhancing personalisation and decoding high-level multimedia components, enhancing engagement and marketing outcomes.

Through the advanced machine learning processes, namely, the random forest regression and the K-Means clustering, this paper pioneers with regard to the literature of social media engagement. Random Forest is more accurate in its predictions and demonstrates the relative significance of the variables like the presence of celebrities and tone of the emotion and visual arrangement, providing greater interpretability compared to the conventional linear models. In addition to this, K-Means clustering reveals concealed patterns of audience-content interaction by identifying Reels that share certain engagement features, thereby making it possible to acquire a multidimensional perspective of user engagement. Combined, these strategies go beyond traditional analytical models and show how data-driven modeling can be able to capture the dynamic and algorithmically mediated character of Instagram Reels.

To further increase the level of analytical rigor, the current study incorporates five theories: Uses and Gratifications Theory (UGT), which assumes that the users interact with the content according to their personal requirements: either entertainment, socializing, or information search (Sichach, 2023:3). Considering the unplanned and entertaining character of Reels, UGT can be used to explain why people choose and consume this content. According to Social Presence Theory, digital cues such as video, voice and facial expression improve intimacy and immediacy both vital elements in drawing viewer attention within a short-term video (Gunawardena, 1975:148).

The Parasocial Interaction Theory can be very relevant to Instagram Reels, particularly the ones with the celebrities and influencers. It examines the one-sided but psychologically influencing closeness users form with figures of the media, which can lead to the further trust and interaction (Horton and Wohl, 1956:220). The Social Identity Theory can recognize how users identify with content that supports or upholds their own values or social identities which will affect the nature of their interactions and attachments to creators or brands (Tajfel, 1974:66). Emotional Contagion Theory explains how the expression of emotions in Reels—such as joy, excitement, or empathy—can elicit corresponding emotional responses in audiences, thereby enhancing engagement (Kramer et al., 2014)

In this regard, the literature review highlights that the analysis of short-form video interaction requires a review of modern theoretical frameworks, as well as, the application of computational approaches. The proposed research fills a significant knowledge gap by bridging machine learning with the existing social media theories and offering practical implications to support marketers in their work and making valuable contributions to the developing body of knowledge on engagement in high-velocity, algorithm-driven content ecosystems.

Methodology

This study was designed to identify the main factors influencing engagement with Instagram Reels, particularly among content published by global celebrities and top brands. The methodology was chosen to align with the study's aim of uncovering engagement patterns through a robust and replicable approach. Both the design and techniques used were selected for their empirical rigor and applicability to real-world digital marketing challenges.

Study Design and Data Collection

Our data sample comprised 438 Instagram Reels, shared in January-March 2024. In order to cover diverse and yet influential content, the research selectively picked five internationally known celebrities and five top brands, according to the rankings in follower numbers extracted out of Statista at the point of data collection. The celebrity accounts of the celebrities, including Cristiano Ronaldo, Lionel Messi, Selena Gomez, Kylie Jenner, and Dwayne Johnson, are selected because they are always ranked among the most-followed people around the world, and they are different categories of entertainment and sports. Nike, National Geographic, NASA, Victoria Secret and ZARA comprise the brand accounts to ensure differentiation in the lifestyle, fashion, science communication and consumer goods industries. Every Instagram Reel published by these accounts within the set three-month period was systematically observed and manually coded. Two educational scholars watched each video on numerous occasions to elicit content-related, emotional, presentation, and demographic characteristics. Before the observation, a structured coding sheet was created with the pre-destined categories of the presence of humans, the way the celebrity looks, the prevailing emotion, thematic framing, setting, type of audio, and visual features. The coders were asked to take notes of their interpretations and cross-validate them using comparison sessions. The discrepancies were discussed and resolved through consensus. In addition, inter-coder reliability was assessed using Cohen's Kappa, yielding a satisfactory level of agreement, which supports the reliability of the coding scheme. This triangulated process of manual-coding was used to guarantee similar classification by qualitative content features prior to machine-learning analysis. The last dataset consisted of both human-rated variables and numerical values (likes, views, comments, engagement rate). These fortified and systematically confirmed observations amounted to a strong analytical base to study the impact of certain content features on the engagement dynamic of Instagram Reels by leading celebrities and brands. However, the sample is limited to five celebrity and five brand accounts over a three-month period, which may restrict the generalizability of the findings across different user groups, content creators, and geographical contexts.

Variables and Operationalization

The rate of engagement is the most significant outcome variable in this investigation. The ratio of the total number of interactions between the viewer and the account and the number of followers of the account is the meaning of the engagement rate. The number of engagements was calculated by adding the likes, comments, shares, and saves of each Reel and then dividing it by the total number of followers at the point of the post, and lastly multiplied by 100. This is because the normalised measure which is presented by this measure only makes it feasible to compare accounts in a meaningful manner even when the size of the audience does differ significantly. While this normalized measure enables comparability across accounts with different audience sizes, it may not fully capture algorithm-driven exposure dynamics. In particular, Reels content can reach audiences beyond followers through the 'Explore' algorithm, which may influence engagement rates independently of follower count. The creation of an ordered set of independent variables was done with an intention to receive the visual, emotional, and contextual features of the material. The binary indicator was used to find out whether the video contained a celebrity or brand in a direct or indirect way, the type of content represented in the Reel was also considered. This was human oriented content, animal oriented images and landscape oriented scenes among others. There was also tagging of the demographic cues, and this allowed them to establish whether the key subjects in the movie were adults, children, or popular figures in the society. After applying systematic observation processes, all of the Reels were grouped into four categories that were used to classify them in terms of the emotional tone of the Reel: lively, positive, neutral, or negative. The extra criteria included the colour scheme in general (vivid, muted, dark, or pastel) and the inclusion or exclusion of sound which recorded the use of audio elements, including music, voiceovers, or ambient sound. The application of one-hot encoding was used to convert all of the categorical characteristics into numerical representations. This was so as to make sure that every category was regarded as an independent variable. Consequently, the data have been made available to machine learning analysis. To avoid scale-related distortions during model training, numerical variables

were z-score normalized. Prior to normalization, extreme values were examined and treated to reduce the influence of outliers on the standardization process. This preprocessing pipeline ensured a coherent and analytically robust dataset suitable for both feature selection and clustering tasks.

The current research is based on a multi-theoretical premise that comprises of five supplementary perspectives as explaining the mechanisms by which Instagram Reels can bring about user engagement. All these theories Uses and Gratifications Theory, Social Presence Theory, Parasocial Interaction Theory, Social Identity Theory, and Emotional Contagion Theory are framed together to explain the effect of various content traits on audience perceptions, motives and exchange. The combination of them offers a consistent analytical prism that rationalizes the choice of variables, directs the discussion of machine-learning results, and enhances the research value of the study to digital marketing and social media engagement studies.

UGT has a direct influence on the choice of independent variables like content type, visual design, shown activity, and sound presence. It is presumed that users will be more exposed to Reels that can generate the highest gratifications such as entertainment (e.g., sports, cityscapes), cognitive stimulation (e.g., creative/intellectual activities), or social value (celebrity appearances). The outcomes of the random forest analysis which reveal the significance of the content created and human-focused elements are in line with the assumptions of UGT that the content created should satisfy the need of the reader to create engagement. This theory is the basis of such variables as human activity, age group, demographics, and audio presence. The importance of human-related variables in the Random Forest model is significant to prove that the higher the social presence expressed in the Reels, the more the engagement can be created, i.e. the more people are used in the Reel (adults, celebrity, demonstration). The findings of the clustering, when the human-centric and expressive clusters generate more interaction, are another confirmation of how social presence can influence the viewer responses. The outcomes of the clustering process when celebrity-led Reels created unique segments of high engagement explain further how parasocial relational cues determine user interaction patterns. The social identity theory is in favour of the addition of demographic and contextual variables (e.g., age group, content themes, and environmental background e.g., human and city). In the same manner, the existing worse results of child-oriented content indicate poorer identity congruency among the average user population on the platform. The variable of emotional tone, being measured on a four point scale (energetic, positive, neutral, negative), is directly derived out of the Emotional contagion theory. The emotion scoring of the importance ranking in the Random Forest and its clustering tendencies show how the emotional resonance influences the user interaction and contributes to the implementation of the theory to the short-format digital content. These theories combined provide a comprehensive prism with which the empirical findings can be interpreted. UGT describes the reason users are interested in a specific kind of content, the Social Presence Theory explains the effects of audiovisual cues on the perceived connectedness, the Parasocial Interaction Theory explains the impact of celebrity-led interaction, the Social Identity Theory explains the impact of demographic and contextual congruence, and the Emotional Contagion Theory captures the effect of affective cues in Reels.

Random Forest Analysis of Feature Importance

In this research, Random Forest has been used as a supervised machine-learning model to determine the most effective predictors of Instagram Reel engagement. Though this study is not aimed at predictive modeling as the ultimate goal, the extracting features of the most important features was carried out through the Random Forest algorithm, which allows getting an information understanding of what content features have the strongest impact on the engagement rate. The method is especially most effective on social-media data that tend to include nonlinear relationships and interaction effects as well as heterogeneous feature scales. The Random Forest model was implemented with 500 trees ($n_{\text{estimators}} = 500$), a maximum depth of 10, and a fixed random state of 42 to ensure reproducibility. Given the exploratory nature of the study, the model was primarily used for feature importance estimation rather than predictive optimization. To assess the robustness of the results, a 5-fold cross-validation procedure was employed, and the stability of feature importance rankings was examined across folds. The consistency of these rankings indicates that the identified key drivers of engagement are stable and not driven by random variation in the data. Model parameters are reported for transparency and replicability. By obtaining an average of the outcomes of all trees, Random Forest can create a strong and consistent ranking of the importance of variables, and the risk of overfitting is minimized, as well as sensitivity to noise, which is very likely in such cases as the analysis of the dataset of multimedia content. A ranked list of features that have an impact on the variation in the engagement rate was produced. Celebrity/brand presence, content type, human-based action, tone of emotion, and demographic are identified as the most significant influencing factors. The values of

importance were based on the average reduction in the prediction loss when each variable was added and this gave a quantitative measure of the relative weight of each characteristic of content. This method therefore facilitated the replicable and clear definition of determinants of engagement without any subjective assumption or manual weighting.

Content Pattern K-Means Clustering

Although Random Forest answered what features were the most significant, unsupervised learning was needed to know how various Reels are organised by nature to form clusters. To this end, K-Means clustering was used in order to identify any latent patterns of content and divide Instagram Reels into homogenous behavior clusters. The algorithm has the following steps: First, a set of KKK cluster centroid is initialized, and each Reel is assigned to the closest centroid by Euclidean distance in the standardized feature space. Once assigned, re-calculations of new centroids are made as the average location of the observations in each cluster. This is repeated until convergence is reached- when no significant changes are seen in assignments anymore. All numeric and the encoded categorical variables were standardized before clustering to make them have equal input to the distance measure. The elbow method identified the number of clusters with the most reduced returns as one of three clusters. These groups reflected different patterns of engagement behavior, including highly engaging human-focused creative posts, less interesting brand-focused posts, and hybrid-content videos with intermediate interaction rates. Two-dimensional visualization of the clustering results was done using Principal Component Analysis (PCA) to increase the interpretability. This move saved the majority of the variability in the multidimensional data besides allowing the readers to see definite clusters divisions as well as content clusters.

Analytical Techniques

To analytically investigate the factors of Instagram Reels engagement, the research project followed a multi-step analytical pipeline, which incorporates exploratory data analysis, feature selection based on machine-learning, and unsupervised clustering. Initially, descriptive and exploratory analyses were performed to get an idea of the fundamental or original composition of the data, distribution of engagement measures, and occurrence of content properties. The stage entailed the analysis of central tendencies, dispersion, correlation patterns, and visual analytics like boxplots and heatmaps to form the initial content-engagement rate relationships. Then, a Random Forest model was utilized, but not in the framework of predictive regression, but rather in terms of the estimations of feature importance, which allowed defining the content attributes that made the biggest contribution to the variability of the engagement rate. Random Forest was chosen due to the strength to non-linear relationships, multicollinearity of variables, and the known capability to work in high-dimensional social media datasets with variable interactions being intricate. The model provided a level of importance of all the video features that were coded, and due to the importance, the study was able to objectively establish the kind of variables that will be prioritized in future analysis. Lastly, K-Means clustering was used to cluster Reels using natural content clusters in terms of their feature compositions and engagement behaviors. Before clustering, numerical variables were standardized to contribute equally to the calculation of the distances and the optimal number of clusters was calculated with the help of the Elbow Method and silhouette scores. This step of unsupervised learning enabled the study to identify underlying content archetypes and patterns of engagement that would not have been identified using the linear analysis or the univariate analysis. The methodology also achieved both interpretability and analytical rigor through a combination of these complementary analytical methods to offer a replicable framework of uncovering content-engagement dynamics in the short-form video platforms.

Model Evaluation and Controls

In this study, the Random Forest model was used as the feature-selection method to perform a systematic selection of content attributes that have the strongest effect on engagement rates on Instagram Reels. The model was used specifically as a predictive regression tool instead of using Rand forest, because the model also provides the relative importance of each feature based on the inherent impurity-based importance measures. Once the data has been preprocessed, and all the categorical variables are coded, the Random Forest algorithm built an ensemble of decision trees all of which are trained on a random sample of the data and features. In constructing the tree, the algorithm chose variables in each node recursively, again and again, that minimized impurity, as defined by a measure of impurity like the Gini impurity or variance reduction. The model combined these reductions in all trees within the forest to produce a normalized importance value of each variable as its total contribution to the variability in engagement

outcomes. The results of the feature-importance can be summarized in Table 1 and obviously show what aspects of Reel content have the most significant impact on engagement. The most powerful driver was the variable of the "Celebrity & Brand presence" which represented 22.10 percent total importance score. This implies that audiences react intensely to recognizability and social identity is revealed in a work involving familiar faces or familiar brands. The other variable that was closely related to the variable Content type was 19.03 meaning that the thematic focus of the video, human-centric scenes, landscapes, and objects is a critical variable in engagement. The demographic cues included in the age group, like adults, children, or mixed-age presence, added 16.27, which means that viewers have a different reaction to the perceived target demographic or the age representation in the content. Equally, Human activity was identified to contribute 16.16, which is a clear indication that dynamic sequences that are action oriented are usually more attractive compared to the non-moving images. Variables ranked below also shed a significant amount of light. The contribution of "Emotion" was 5.99 indicating that although emotional tone could play a role in performance, it is still subordinate to other more fundamental factors i.e. identity cues and content structure. Lastly, the value of the importance of "Color" was very low (0.16%), meaning that aesthetic features such as color scheme are of little importance when other contents features are included. In general, the Random Forest analysis allowed the study to identify and measure the most significant driving factors of engagement data-drivenly, which will serve as a strong basis in further clustering and interpretation.

A summary table of the feature importance percentages provided additional clarity, listing the key drivers of engagement as follows:

Table 1. Key Drivers of Engagement

<i>Feature</i>	<i>Importance (%)</i>
C&B	22.10
Content	19.03
Age Group	16.27
Human Activity	16.16
Emotion	5.99
Color	0.16

Results

The study findings are consistently in support of the major findings of the study. The detection of celebrity and brand presence (C&B), and content composition as the major engagement drivers can be logically related to the theoretical framework that has been discussed above. These findings are again confirmed by the clustering results which indicate different profiles of engagement among categories of content. The minor inconsistencies- the unsuspected low engagement of positive emotional tone- are mentioned in discussion section, which provides the plausible interpretations with literature-backed explanations. No overinterpretation or unnecessary claims have been made. The figures and tables that have been shown support all the critical points of data that have been mentioned.

Random Forest has not been applied in this paper as a predictive regression model but as a feature importance estimator to identify the most influential variables on the engagement rate. The algorithm was used in supervised-learning situation where the engagement rate acted as the target variable, and all the content-related attributes became the predictors. Random Forest will be of great help in this case since it builds a collection of decision trees that are trained on a random sample of the data and features. The model uses the repeated estimation of the amount of impurity reduction by each feature in hundreds of trees to produce a stable and reliable measure of feature importance. This method is already established in determining important drivers in nonlinear and complex datasets, that is, when the variables are categorical as well as continuous. The model that was used in this study, the Random Forest, has been optimized to be standard, which guarantees the stability and interpretability to be balanced. The importance of the features was determined by averaging the reduction in the mean squared error (MSE) of all trees in which the feature was used to split. The given methodology offers a clear and repeatable ranking of engagement factors.

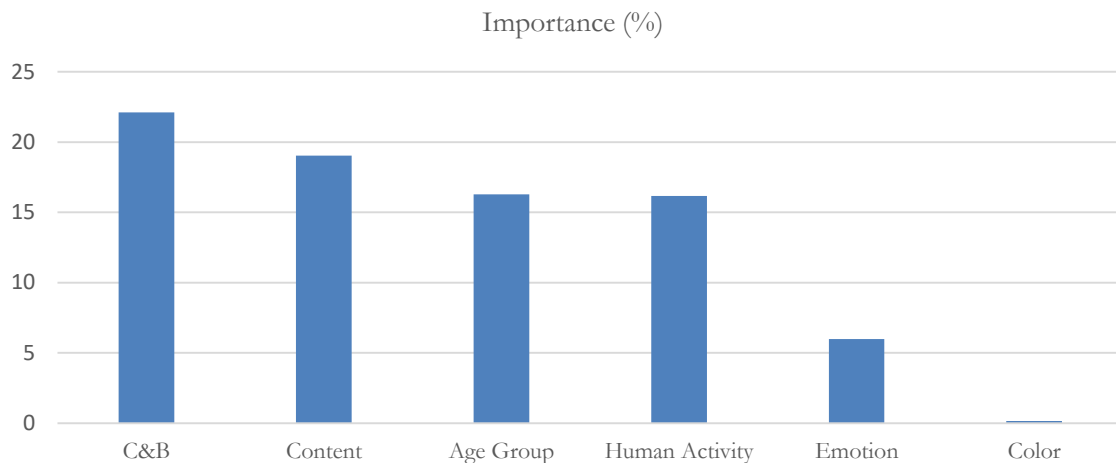


Figure 1. *Key Drivers of Engagement*

The results of the Random Forest indicate that the most influential factor when it comes to engagement is the celebrity/brand presence (22.10). This shows that the emergence of a familiar character or brand that is trusted adds great value to the user engagement, which is in line with the theories of parasocial interaction and the effectiveness of celebrity endorsing. Other significant factors were content type (19.03%), age group represented (16.27%) and depicted human activity (16.16%). This evidence indicates that the specifics of the visual narrative, including the presence of adult and celebrities in the video, human-city pairings, and sporting events, are the key factors in determining the way viewers react to the short video content. The somewhat small values given to emotion (5.99%) and color (0.16) mean that these factors are relevant to user attention, but secondary to demographic and contextual factors in the video. Notably, the low contribution of the color facilitates the empirical finding of the study that the aesthetic variables alone cannot be used to drive engagement without meaningful human-centered content. Combined with other findings, these results determine a distinct sequence of engagement driving factors and give an evidence-based recommendation of how to optimize content on Instagram Reels.

Clustering Analysis

To continue comparing the patterns in the underlying content that were not covered by the supervised analysis, the K-Means method was used as a form of unsupervised analysis. Although Random Forest shows the most important features, the clustering will show how the various combinations of features naturally cluster together and whether they exhibit different engagement behaviors. The K-Means algorithm divides the observations into clusters through minimization of the variance within a cluster. All numerical variables were standardized and categorical variables one-hot coded before clustering. The most desirable k was obtained by using the elbow method, which analyzed the pace of enhancement of within-cluster sum of squares. To facilitate interpretation, Principal Component Analysis (PCA) was used to minimize the multidimensional features space into two major components that explain most of variance. This not only allowed the principles to be visualized and make meaningful interpretations of categories of content qualitatively.

Cluster 0 had reels that were rich with expressive or narrative human activity sports, performances, creative acts, and human interactions that were city based. A common pattern in these videos was the highest level of engagement which proves the relevance of relatability, motion and dynamic human representation. Cluster 1 was largely brand-focused reels that had little emotional appeals or geographical differences. The engagement in this cluster was very low throughout, which supports the notion that purely promotional material, devoid of the human factors, does not work well in the platforms that rely on the sense of personal connection and entertainment. Cluster 2 consisted of hybrids (human + animal, product + environment). The levels of engagement were average and less predictable, which means that mixed-content structures are attention-grabbing, but they are not as emotion- or narrative-driven as Cluster 0. The distinct division of clusters underlines the importance of feature-based analysis in addition to unsupervised learning. Whereas the most influential content attributes are identified by Random Forest, clustering can also give a contextual information on how these content attributes co-exist and how users react to the patterns.

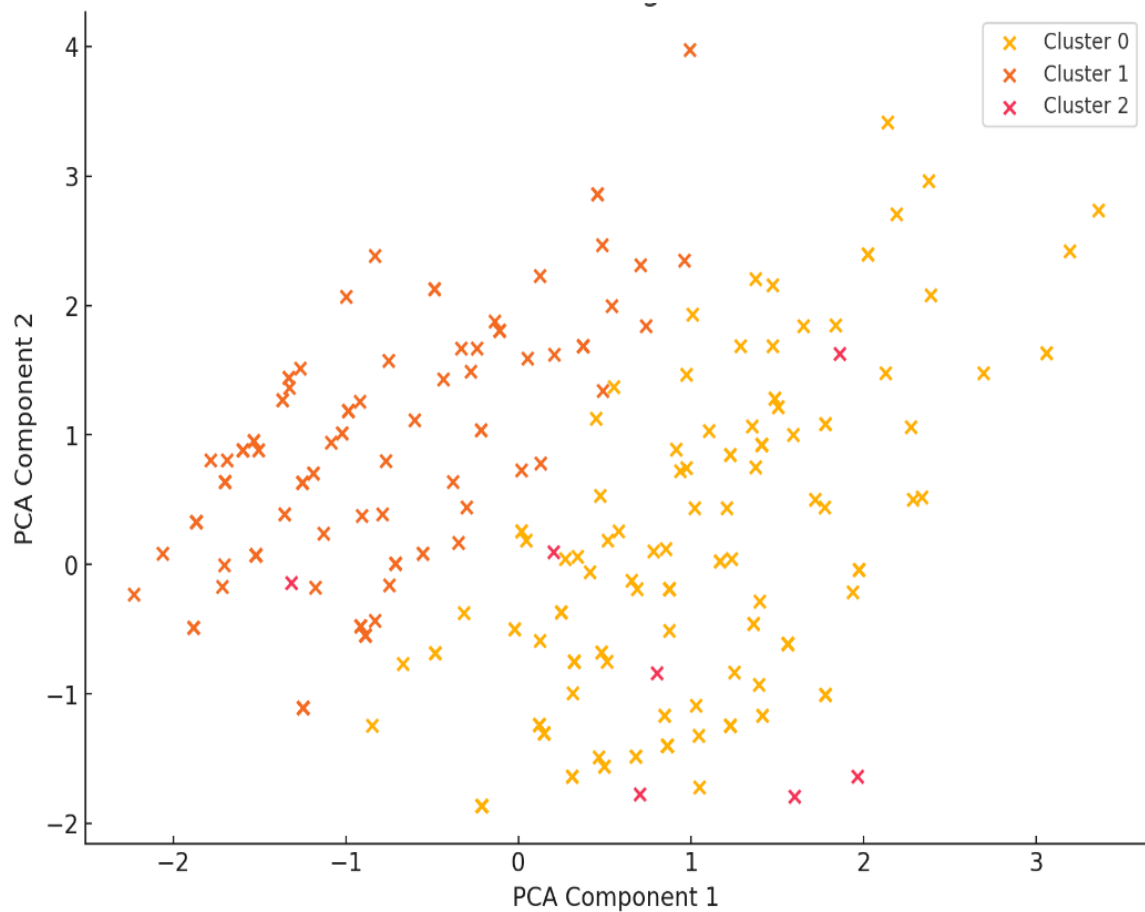


Figure 2. *K-Means Clustering Visualization*

Here are the summarized engagement rate results:

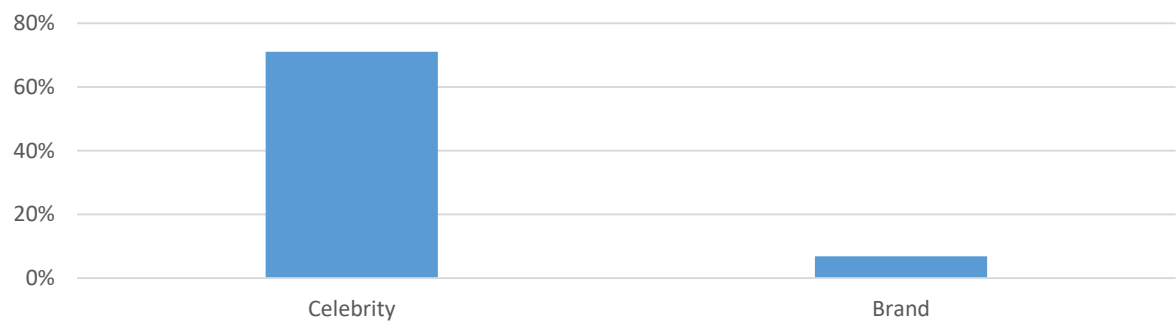


Figure 3. *Impact of Celebrity vs. Brand on Engagement Rate*



A comparative evaluation of engagement rates for content featuring celebrities versus content focused solely on brands is presented in Figure 3. The results show a substantial difference in performance: Reels featuring celebrities achieved an average engagement rate of 6.84%, whereas brand-focused Reels yielded a considerably lower average engagement rate of 1.71%

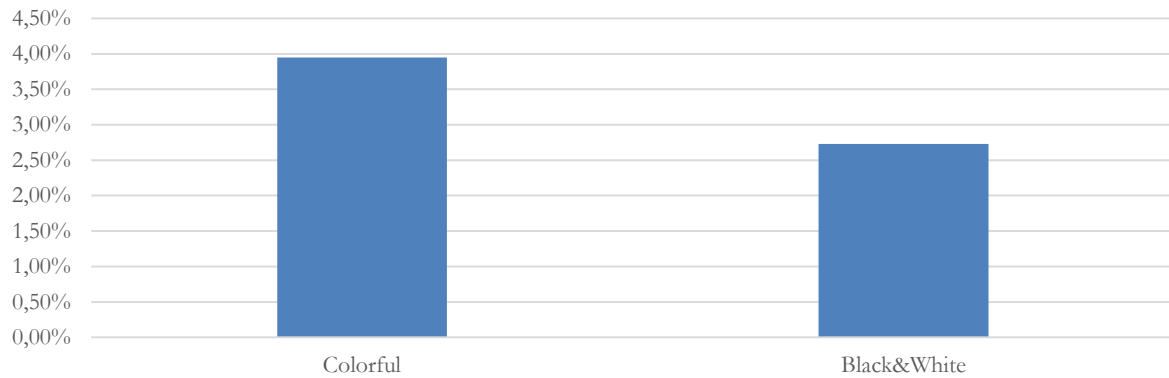


Figure 4. *Impact of Color on Engagement Rate*



Figure 4 displays the comparative effect of colour schemes—in particular colourful as opposed to black-and-white visuals—on the engagement price of Instagram Reels. The consequences clearly suggest that colourful movies considerably outperform black-and-white content in terms of audience interaction. Colorful movies finished a median engagement charge of 3.95%, while black-and-white movies generated a decrease average engagement rate of 2.73%.

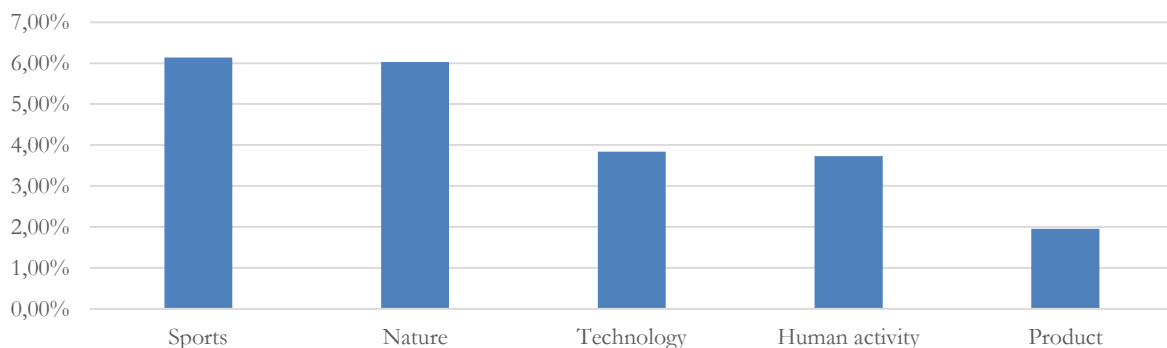
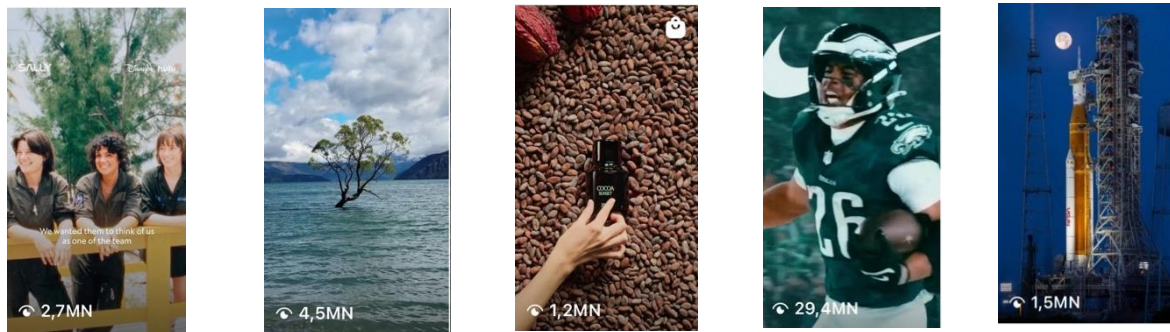


Figure 5. *Impact of Activity/Post Types on Engagement Rate*



The investigation provided in Figure five is concerned with figuring out the volume to which the engagement costs of Instagram Reels are prompted by using the numerous types of interest and posts which might be presented at the platform. According to the facts that became analysed, content material this is related to sports (6.14%) and content this is focused on nature (6.03%) are the two categories that produce the very best engagement rates across all of the classes that have been analysed. On the alternative hand, the least effective approach of catching user attention is the advertising of products, which has an engagement charge of 1.96%. At the identical time, there's a center floor occupied by content that relates to era and innovation (3.84%) as well as widespread human hobby (3.73%), which indicates that the ranges of contact are slight.

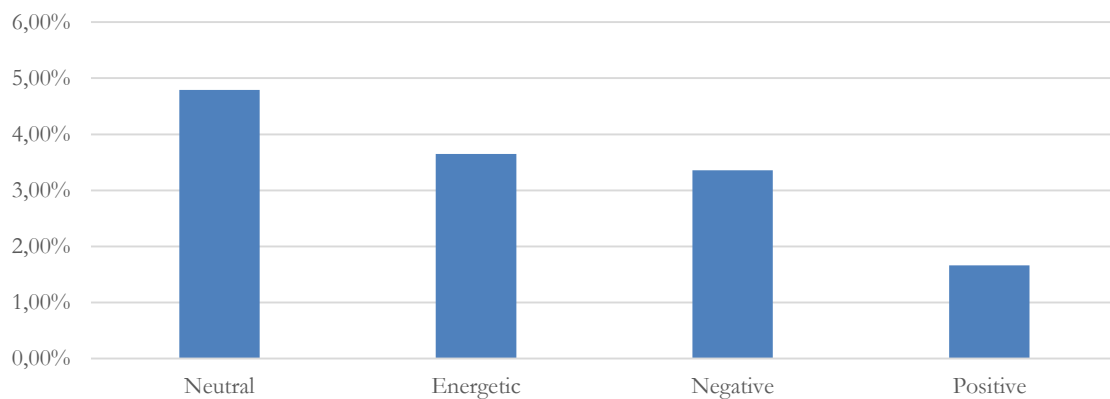
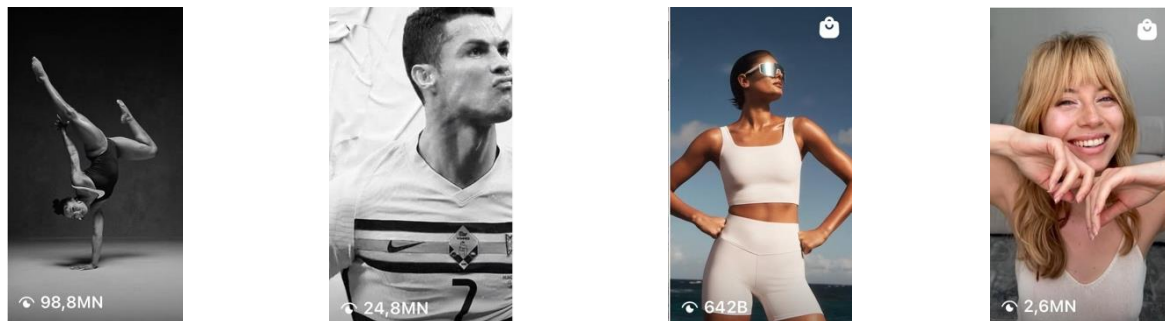


Figure 6. *Impact of Emotion on Engagement Rate*



The impact of emotional tone on the engagement costs of content discovered in Instagram Reels is depicted in Figure 6. It is interesting to word that the records reveals that content material that is impartial in tone receives the highest engagement charge of 4.79%, that is greater than that of some other emotional category. This discovery calls into query the commonly held perception that emotionally charged content, and mainly the ones that are uplifting or full of power, is more likely to bring about extra levels of interplay. The following categories, lively (3.65%) and poor (3.36%), show a modest stage of participation. It is surprising that fabric with a pleasant tone, that's normally idea to carry out the satisfactory in digital advertising, had the lowest interaction charge of most effective 1.66%. The reality that this outcome was now not expected shows that visitors can be more receptive to authenticity or subtlety in assessment to overt emotional appeals, or it is able to be the case that information this is excessively nice could be seen as inauthentic or business. The diffused function that emotional framing performs in the improvement of a digital content approach is introduced to mild by using these findings. To preserve the eye of viewers and encourage meaningful connection, entrepreneurs and content material creators may want to rethink preconceptions regarding emotional enchantment and as an alternative explore more diffused, available tones.

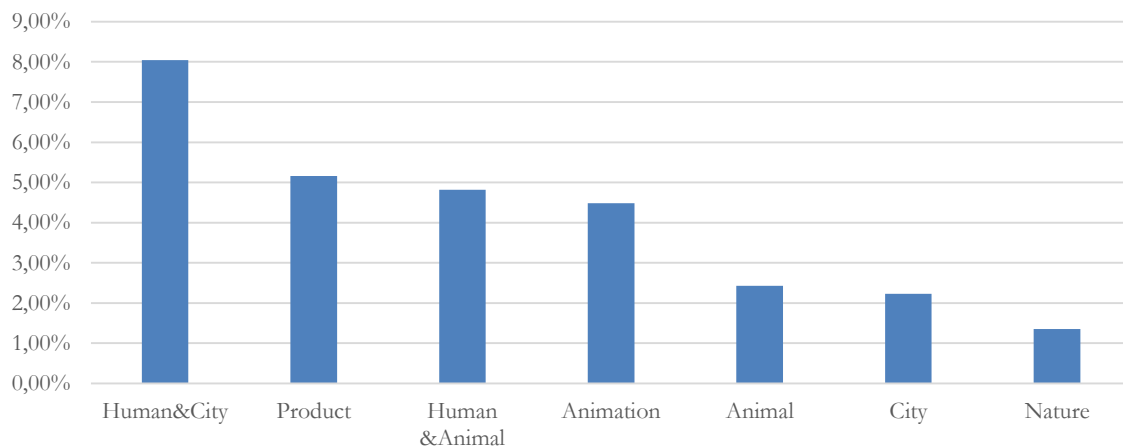


Figure 7. *Impact of Content Types on Engagement Rate*

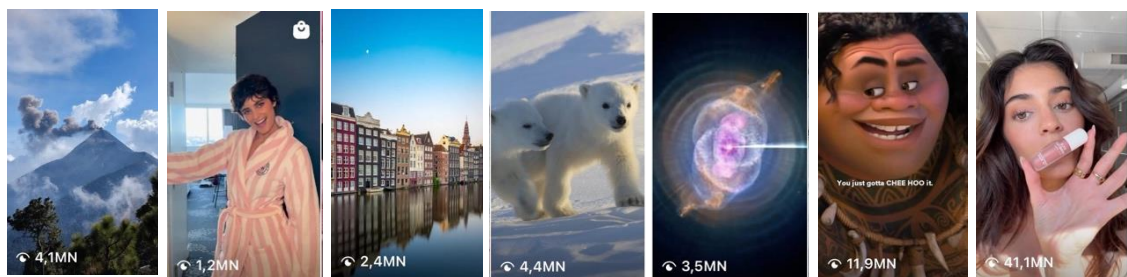


Figure 7 shows a close analysis of the participation rates of different types of content in Instagram Reels. The distribution of the analysis points to the fact that the most engagement can be attained in the case of the use of human presence mixed with the urban setting, as it was observed in the Human and City category, where the average engagement rate was 8.04%. This indicates that viewers are especially sensitive to the material which combines identifiable human factors and dynamic street scenes. Content about products did also work, as it achieved the 5.16% engagement rate, with Human and Animal combinations (4.82%) and animated content (4.48) being close. The implication of these findings is that the content that is satisfying in terms of the storytelling potential or emotional relatability, particularly when it comes to humans, animals, or imaginative visuals, is more appealing to the viewers. On the other hand, nature-only content had the least level of engagement of 1.35, although it is popular with visual channels. Likewise, independent City (2.23%) and Animal (2.43) content performed under-average, showing that in isolation, isolated features are not as compelling as they are when they interact with humanism or context.

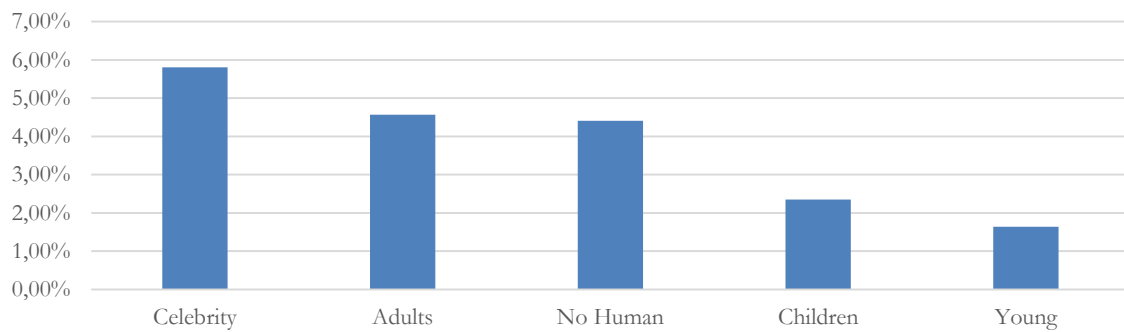


Figure 8. *Impact of Age Group on Engagement Rate*



As Figure 8 shows, various representations created by age in Instagram Reels content affect the engagement rates differently. The evaluation shows that celebrity appearances have the best engagement rate at 5.81 which highlights the fact that familiar personalities will continue to be popular and effective in attracting interaction among the audience. It is then succeeded by material of adults (4.57%) and no human being whatsoever (4.41%), both of which perform fairly well, which implies that familiarity (adults) or visual narrative with no human being can keep the viewers focused. Conversely, the content that contained children (2.35) and younger people (1.64) had the fewest views on content, meaning that these types of images may not attract the attention of the wider Instagram audience as well. It is especially interesting considering that the audience of this platform is younger in age, and it is possible that the viewers would better react to aspirational or professional figures (celebrities or adults) and not their peers or younger models.

Discussion

This section interprets the empirical findings in light of the theoretical framework and prior literature on social media engagement. Celebrity-driven content appears to have a stronger effect on attracting user attention and encouraging interaction. The findings suggest that after viewers have a feel of familiarity with or a private connection to a prominent discern, they're much more likely to interact with the content that they're viewing. This tendency may be a result of the psychological advantages of parasocial interplay, as well as the emotional resonance that celebrities regularly generate. On the opposite hand, content material this is targeted around a brand, which may not possess the equal diploma of private attraction or emotional narrative, tends to create a weaker reaction from users. The importance of employing superstar endorsement processes within the promotion of brief-form virtual content is in addition substantiated via these discoveries. This finding is consistent with Parasocial Interaction Theory, which suggests that perceived intimacy with public figures enhances user engagement.

Vibrant, visually attractive content material captures user interest greater successfully, aligning with previous findings in visible communication and social media aesthetics. The preference for colorful content

possibly stems from its potential to rouse more potent emotional responses and beautify the perceived enjoyment cost of brief-form motion pictures. These effects underscore the significance of visual design alternatives in social media marketing techniques, particularly on systems like Instagram in which aesthetic fine performs a crucial function in content overall performance. This aligns with prior literature emphasizing the role of visual salience in attracting user attention (Kietzmann et al., 2011).

Dynamic and immersive snap shots, which include those that depict athletic performance or natural landscapes, are specifically nicely obtained by using audiences once they inspire emotions of excitement or different emotions. These findings are consistent with studies highlighting the role of emotional and narrative content in driving engagement (Berger & Milkman, 2012).

The findings spotlight the significance of narrative and aesthetic cost with regards to content material this is digital in nature. Even if useful or promotional postings may serve strategic enterprise motives, it appears that evidently fabric that both entertains users (for example, sports) or conjures up them (for example, nature) generates a better level of engagement from customers. This emphasises the need of content material makers and marketers going past simply communicating product data and instead growing immersive reviews which can be in keeping with the emotions and pastimes of the stop user. From a theoretical perspective, this may be explained by Emotional Contagion Theory, where subtle or neutral emotional cues may be perceived as more authentic and less commercially driven, thereby fostering stronger user responses.

Findings highlight the importance of context-based content, which utilizes the human element and the storytelling approach. Marketers who want to maximize their engagement when on Instagram should think of using image-based attractive backgrounds (e.g., city or animated scenes) and a believable human experience to ensure that the views and interactions get as many as possible. This finding can also be interpreted through Social Presence Theory, as human-centered and context-rich content enhances perceived closeness and interaction.

This research demonstrates the significance of a demographic representation in visual strategy, particularly to brands or creators who want to maximize reach and engagement. The inclusion of adult or celebrity characters or even the creation of content with interesting stories that do not involve human beings can be more likely to act as an effective way of engagement than content created or directed to or involving children and teenagers. A comparative study of the high and low engagement Instagram Reels showed that there are differences in the content patterns. The most actively involving videos were usually ones that involved celebrities, involved creative or intellectual action, and had an emotionally evocative tone- especially one of neutrality or excitement. Low-engagement videos on the other hand contained a larger proportion of brand-oriented video content that lacked a significant amount of emotional range which implied that the homogeneity of content could be an obstacle to user engagement and retention. This pattern is consistent with Social Identity Theory, suggesting that users engage more with content that reflects aspirational or socially valued identities.

One of the most notable findings of the study is that neutral emotional tone generated higher engagement than explicitly positive or lively content. This result appears to diverge from conventional interpretations of Emotional Contagion Theory, which generally assume that emotionally charged content stimulates stronger audience responses. A possible explanation is that users in short-form video environments may experience fatigue toward highly stylized or overtly promotional emotional appeals. In this context, neutral-toned content may be perceived as more authentic, less strategic, and therefore more credible. This interpretation is particularly relevant in algorithm-driven environments such as Instagram Reels, where users are exposed to a high volume of emotionally intensified content and may respond more positively to subtle, less performative forms of expression.

The findings both align with and extend prior literature. While previous studies have emphasized content type, emotional tone, and visual appeal as key engagement drivers (Alhabash & Ma, 2017; Voorveld et al., 2018), the present study demonstrates that identity-related cues, particularly celebrity presence, play a more dominant role in short-form video environments. Furthermore, unlike prior research that examines these factors in isolation, this study shows how multiple content attributes interact and form distinct engagement patterns.

Conclusion

This paper provides a quantitative and analytical examination of the factors influencing user interaction on Instagram Reels, focusing on content produced by globally recognized celebrities and major brands. Through a machine-learning-enhanced methodological design, the study offers insights into the role of visual, emotional, and demographic factors in shaping engagement within short-form video environments. The findings indicate that celebrity presence, human-centered content, demographic representation, and emotional tone are among the most influential drivers of engagement. These results are broadly consistent with theoretical perspectives such as Parasocial Interaction Theory, Social Presence Theory, and Emotional Contagion, highlighting the importance of relational and affective cues in digital content consumption.

Methodologically, the study contributes to digital marketing analytics by integrating feature importance estimation using Random Forest with content clustering through K-Means. This combined approach enables the identification of both key engagement drivers and underlying content patterns, offering a transparent and replicable analytical framework for examining algorithm-driven content performance in social media contexts.

From a practical perspective, the findings provide actionable implications for marketers and content creators. In particular, the use of familiar or recognizable figures, human-centered narratives, and context-rich visual storytelling appears to be associated with higher engagement levels. Moreover, the results suggest that emotionally neutral tones may, in certain contexts, generate stronger user responses, indicating the need for a more nuanced understanding of emotional expression in short-form video strategies. These insights encourage a shift from purely promotional content toward more narrative-driven and audience-oriented formats.

Despite these contributions, the study has several limitations. The dataset is limited to a three-month period and a selection of high-profile accounts, which may constrain the generalizability of the findings to smaller creators or different cultural contexts. Future research could extend the temporal scope, incorporate more diverse account types, and conduct cross-platform comparisons with platforms such as TikTok and YouTube Shorts. Additionally, incorporating longitudinal or experimental designs would help to establish causal relationships between content attributes and engagement outcomes, while further exploration of algorithmic visibility mechanisms could deepen understanding of platform-specific dynamics.

In sum, this study contributes to bridging the gap between theoretical perspectives and data-driven approaches in digital marketing by demonstrating how machine learning techniques can be used to systematically analyze engagement in short-form video content. By offering a scalable and empirically grounded framework, the research provides a foundation for future studies on the evolving dynamics of social media engagement.

Ethical Declaration

The study entitled “*Instagram Reels Engagement Drivers: Machine Learning Insights into Celebrities and Brands*” was conducted in accordance with scientific research principles, ethical standards, and citation guidelines. No manipulation, fabrication, or falsification was performed on the collected data. Furthermore, this study has not been submitted to any other academic publication venue for evaluation. Ethical committee approval was not required for this research.

Etik Beyan

“*Instagram Reels Etkileşimini Artıran Faktörler: Ünlüler ve Markalar Hakkında Makine Öğrenmesi İlgörülleri*” başlıklı çalışmanın yazım sürecinde bilimsel kurallara, etik ve alıntı kurallarına uyulmuş; toplanan veriler üzerinde herhangi bir tahrifat yapılmamış ve bu çalışma herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiştir. Bu araştırma da etik kurul kararı zorunluluğu bulunmamaktadır.

Statement of Contribution Rate of Researchers

The first author contributed 60%, and the second author contributed 40%.

Araştırmacıların Katkı Oranı Beyanı

1.yazar: % 60, ikinci yazar % 40 oranında katkı sağlamıştır.

Declaration of Conflict

There are no potential conflicts of interest in this study.

Çatışma Beyanı

Çalışmada herhangi bir potansiyel çıkar çatışması söz konusu değildir.

Funding

No funding was received from any institution or organization for this study.

Finansman

Bu çalışma için herhangi bir kurum veya kuruluřtan destek alınmamıştır.

Statement of Use for Artificial Intelligence and Its Types

Artificial intelligence and its various types were not used in the writing of this article.

Yapay Zeka ve Türleri Kullanım Beyanı

Bu makalenin yazımında yapay zeka ve türleri kullanılmamıştır.

Data Sharing Statement

We declare that, upon reasonable request for the purpose of verifying the findings, we can share the data of this study according to the conditions specified in the relevant section of the "ethical principles and publication policy".

Veri Paylaşım Beyanı

Bu çalışmanın verilerini bulguların doğrulanması amacıyla makul bir talep üzerine “etik ilkeler ve yayın politikası”nın ilgili kısmında belirtilen şartlara göre paylaşılabileceğimizi beyan ederiz.

Notes

This article was presented as a poster at the European Academy of Marketing (EMAC) Spring Conference 2025, held in Pozuelo, Spain, from 27-30 May 2025.

Not

Bu makale, 27-30 Mayıs 2025 tarihleri arasında İspanya'nın Pozuelo kentinde düzenlenen Avrupa Pazarlama Akademisi (EMAC) Bahar Konferansı 2025'te poster olarak sunulmuştur.

References

- Aldhamiri, A., Carlson, J., Vilches-Montero, S., Rahman, S. M., & Gudergan, S. P. (2024). What drives higher active customer engagement in luxury brands' social media? Measurement and contingencies. *Journal of Retailing and Consumer Services*, 79, 103804.
- Alhabash, S., & Ma, M. (2017). A tale of four platforms: Motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students? *Social Media + Society*, 3(1), 2056305117691544. <https://doi.org/10.1177/2056305117691544>
- Al Kurdi, B., Alshurideh, M. T., Hamadneh, S., Kalra, D., Marwaha, S., & Alzoubi, H. M. (2024, February). Machine learning applications in social media marketing: innovating customer engagement and brand building. In 2024 2nd international conference on cyber resilience (ICCR) (pp. 1-7). IEEE.
- Andariesta, D. T., & Wasesa, M. (2023). Machine learning models to predict the engagement level of Twitter posts: Indonesian e-commerce case study. *Procedia Computer Science*, 227, 823-832.
- Asan, K., & Zwiendelaar, J. (2025). Digital leisure use and psychological well-being among recreation management students: The mediating role of social media engagement. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 37, 100570.
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192–205. <https://doi.org/10.1509/jmr.10.0353>
- De Vries, L., Gensler, S., & Leeftang, P. S. H. (2012). Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing*, 26(2), 83–91.
- Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International journal of educational telecommunications*, 1(2), 147-166.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

- Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry*, 19(3), 215–229. <https://doi.org/10.1080/00332747.1956.11023049>
- Jin, X. L., Yin, M., Zhou, Z., & Yu, X. (2021). The differential effects of trusting beliefs on social media users' willingness to adopt and share health knowledge. *Information Processing & Management*, 58(1), 102413.
- Johnson, B. K., & Ranzini, G. (2018). Click here to look clever: Self-presentation via selective sharing of music and film on social media. *Computers in Human Behavior*, 82, 148-158.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, 54(3), 241-251.
- Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790.
- Krisprimandoyo, D. A., Sufa, S. A., Wardani, D. T., & Widiyanto, S. (2024). Exploring the relationship between social media engagement, customer reviews, and brand perceptions: A comprehensive study in retail industry. *International Journal of Business, Law, and Education*, 5(2), 1584-1591.
- Liadeli, G., Sotgiu, F., & Verleghe, P. W. (2023). A meta-analysis of the effects of brands' owned social media on social media engagement and sales. *Journal of Marketing*, 87(3), 406-427.
- Lipsman, A., Mudd, G., Rich, M., & Bruich, S. (2012). The power of "like": How brands reach (and influence) fans through social-media marketing. *Journal of Advertising Research*, 52(1), 40–52.
- Lou, C., & Xie, Q. (2021). Something social, something entertaining? How digital content marketing augments consumer experience and brand loyalty. *International Journal of Advertising*, 40(3), 376-402.
- Meng, K. S., & Leung, L. (2021). Factors influencing TikTok engagement behaviors in China: An examination of gratifications sought, narcissism, and the Big Five personality traits. *Telecommunications Policy*, 45(7), 102172.
- Menon, D. (2022). Factors influencing Instagram Reels usage behaviours: An examination of motives, contextual age and narcissism. *Telematics and Informatics Reports*, 5, 100007.
- Moran, G., Muzellec, L., & Johnson, D. (2020). Message content features and social media engagement: Evidence from the media industry. *Journal of Product & Brand Management*, 29(5), 533–545.
- Omar, B., & Dequan, W. (2020). Watch, share or create: The influence of personality traits and user motivation on TikTok mobile video usage. *International Association of Online Engineering*, 121-137.
- Onofrei, G., Filieri, R., & Kennedy, L. (2022). Social media interactions, purchase intention, and behavioural engagement: The mediating role of source and content factors. *Journal of Business Research*, 142, 100-112.
- Qin, Y. S. (2020). Fostering brand–consumer interactions in social media: the role of social media uses and gratifications. *Journal of Research in Interactive Marketing*, 14(3), 337-354.
- Reddy, D., Singh, A., Chopra, R., & Patel, R. (2021). Leveraging machine learning algorithms and natural language processing for AI-Enhanced social media marketing analytics. *Journal of AI ML Research*, 10(8).
- Reimer, T. (2023). Environmental factors to maximize social media engagement: A comprehensive framework. *Journal of Retailing and Consumer Services*, 75, 103458.
- Schivinski, B. (2021). Eliciting brand-related social media engagement: A conditional inference tree framework. *Journal of Business Research*, 130, 594-602.
- Shao, G. (2009). Understanding the appeal of user-generated media: a uses and gratification perspective. *Internet research*, 19(1), 7-25.
- Sichach, M. (2023). Uses and Gratifications Theory-Background, History and Limitations. *History and Limitations* (November 17, 2023).
- Tafesse, W., & Wien, A. (2018). Using message strategy to drive consumer behavioral engagement on social media. *Journal of Consumer Marketing*, 35(3), 241-253.
- Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information*, 13(2), 65-93.
- Trunfio, M., & Rossi, S. (2021). Conceptualising and measuring social media engagement: A systematic literature review. *Italian Journal of Marketing*, 2021(3), 267-292.
- Voorveld, H. A., Van Noort, G., Muntinga, D. G., & Bronner, F. (2018). Engagement with social media and social media advertising: The differentiating role of platform type. *Journal of advertising*, 47(1), 38-54.
- Wallach, K. A., Pham, H., Koschmann, A., & Arwade, G. (2025). Analyzing the impact of faces on consumer engagement in social media videos: a machine learning approach. *Journal of Consumer Marketing*, 42(3), 318-335.
- Wibowo, A., Chen, S. C., Wiangin, U., Ma, Y., & Ruangkanjanases, A. (2020). Customer behavior as an outcome of social media marketing: The role of social media marketing activity and customer experience. *Sustainability*, 13(1), 189.
- Wies, S., Bleier, A., & Edeling, A. (2023). Finding goldilocks influencers: How follower count drives social media engagement. *Journal of Marketing*, 87(3), 383-405.

GENİŞ ÖZET

Bu çalışma, dijital pazarlama bağlamında giderek daha fazla önem kazanan kısa video formatı Instagram Reels'te kullanıcı etkileşimini (engagement) artıran unsurları veri odaklı bir yaklaşımla belirlemeyi

amaçlamaktadır. Reels’in algoritmik olarak önceliklendirilmesi, tam ekran dikey izleme deneyimi, hızlı tüketim döngüsü ve popüler ses/akım kullanımını teşvik eden yapısı; içerik üreticileri, ünlüler ve markalar için stratejik bir etkileşim alanı yaratmıştır. Buna rağmen, Reels içeriklerinde etkileşimi sistematik biçimde yükselten özelliklerin hangileri olduğu ve bu özelliklerin görece önem sıralaması konusunda ampirik çalışmaların sınırlı kaldığı görülmektedir. Mevcut araştırma, bu boşluğu gidermek üzere makine öğrenmesi temelli yöntemleri geleneksel etkileşim kuramlarıyla birleştirerek, Reels performansını açıklayan içerik özniteliklerini hem “en güçlü belirleyiciler” hem de “içerik örüntüleri” düzeyinde ortaya koymaktadır. Araştırmanın veri seti, Ocak–Mart 2024 döneminde paylaşılan toplam 438 Instagram Reels videosundan oluşmaktadır. Örnekleme, küresel ölçekte yüksek takipçi sayısına sahip beş ünlü (Cristiano Ronaldo, Lionel Messi, Selena Gomez, Kylie Jenner ve Dwayne Johnson) ile beş büyük markanın (Nike, National Geographic, NASA, Victoria’s Secret ve Zara) resmi hesaplarından seçilmiştir. Bu seçim, hem ünlü temelli içeriklerin hem de marka temelli içeriklerin aynı analitik çerçevede karşılaştırılmasına olanak tanıyarak, kısa video ekosisteminde “tanınmış figür etkisi” ve “kurumsal içerik etkisi” ayrımını görünür kılmaktadır. Videolar, yapılandırılmış bir kodlama formu üzerinden iki araştırmacı tarafından birden çok kez izlenerek içerik ve sunum özellikleri açısından etiketlenmiştir. Kodlama; insan varlığı (insan odaklı, insan+şehir, insan+hayvan vb.), ünlü/marka görünürlüğü, içerik türü (spor, doğa, ürün tanıtımı, teknoloji/yenilik gibi), insan faaliyeti (dinamik eylem, performans, yaratıcı etkinlik vb.), duygusal ton (canlı, olumlu, nötr, olumsuz), demografik ipuçları (yetişkin, çocuk/genç, ünlü, insan yok) ve estetik/ses bileşenleri (renk şeması, ses varlığı) gibi boyutları kapsamaktadır. Kodlayıcılar arası tutarlılık, görüş ayrılıklarının uzlaşısıyla giderilmesiyle güçlendirilmiştir. Çalışmanın temel bağımlı değişkeni etkileşim oranıdır. Etkileşim oranı, beğeni, yorum, paylaşım ve kaydetme gibi etkileşimlerin toplamının takipçi sayısına oranlanması ve yüzdelik biçime dönüştürülmesiyle hesaplanmıştır. Böylece farklı takipçi büyüklüklerine sahip hesaplar arasında adil karşılaştırma yapılması hedeflenmiştir. Analitik süreç üç aşamada yürütülmüştür. İlk aşamada keşifsel veri analizi ile dağılımlar, temel eğilimler ve olası ilişkiler incelenmiştir. İkinci aşamada, içerik özelliklerinin etkileşim oranını açıklamadaki görece önemini belirlemek üzere Random Forest regresyon modeli kullanılmıştır. Random Forest’in tercih edilme nedeni, sosyal medya verilerinde sıklıkla görülen doğrusal olmayan ilişkileri ve karmaşık etkileşimleri yakalamadaki başarısı, yüksek boyutlu kategorik değişkenleri (one-hot kodlama sonrası) yönetebilmesi ve değişken önemini yorumlanabilir biçimde sunabilmesidir. Üçüncü aşamada ise denetimsiz öğrenme yaklaşımıyla K-Means kümeleme analizi uygulanmış; Random Forest ile öne çıkan özniteliklerin birlikte oluşturduğu içerik örüntüleri belirlenmiştir. Küme sayısı, dirsek yöntemi ve uygunluk değerlendirmeleriyle üç küme olarak seçilmiş; yorumlanabilirliği artırmak için PCA görselleştirmesiyle kümelerin ayrışması desteklenmiştir. Bulgular, Reels etkileşimini en güçlü biçimde etkileyen değişkenin “Ünlü & Marka varlığı” olduğunu göstermektedir (önem ağırlığı yaklaşık %22.10). Bunu içerik türü (%19.03), yaş grubu/demografik ipuçları (%16.27) ve insan faaliyeti (%16.16) izlemektedir. Duygusal tonun katkısı daha sınırlı düzeyde kalmakla birlikte (%5.99), etkileşim üzerinde yine de anlamlı bir rolü vardır; renk şemasının görece etkisi ise çok düşük görülmektedir (yaklaşık %0.16). Bu sonuçlar, kısa video performansında “salt estetik düzenleme” yerine “insan merkezli anlatı”, “tanınmış figür/kimlik ipucu” ve “dinamik eylem” gibi anlam yüklü bileşenlerin belirleyici olduğunu göstermektedir. Ayrıca ünlü içeren içeriklerin ortalama etkileşim oranının marka odaklı içeriklere kıyasla belirgin şekilde daha yüksek olduğu görülmüştür (örneğin ünlü ağırlıklı içeriklerde ortalama %6.84; marka odaklı içeriklerde %1.71). Bu fark, izleyicilerin tanıdık figürlerle kurduğu parasosyal bağın etkileşimi tetiklediğini ileri süren yaklaşımla uyumludur. Kümeleme analizi, içeriklerin üç temel örüntü altında toplandığını ortaya koymaktadır. Birinci küme, spor/performans, yaratıcı etkinlikler ve insan+şehir gibi “anlatı taşıyan, hareketli ve insan odaklı” içeriklerden oluşmakta ve en yüksek etkileşim düzeyini üretmektedir. Bu kümede, izleyicinin kendini içerikle özdeşleştirmesini kolaylaştıran sosyal bulunuşluk ipuçları (yüz, beden dili, etkileşim, hız ve dinamizm) daha belirgin görülmektedir. İkinci küme, daha çok “marka/ürün merkezli, duygusal anlatısı zayıf ve tanıtım ağırlıklı” içeriklerden oluşmakta ve en düşük etkileşime sahiptir; bu bulgu, kısa video ortamında yalnızca ürün odaklı mesajların izleyicide “bağ kurma” ve “katılım” üretmekte yetersiz kalabildiğine işaret etmektedir. Üçüncü küme ise insan, ürün, hayvan ve çevresel unsurların birlikte kullanıldığı “hibrit” içeriklerden oluşmakta; etkileşim düzeyi orta ve daha değişken bir dağılım sergilemektedir. Bu küme, farklı öznitelik kombinasyonlarının belirli koşullarda işe yarayabildiğini; ancak en güçlü performansın genellikle insan ve kimlik ipuçlarını merkeze alan içeriklerde toplandığını desteklemektedir. İçerik türleri karşılaştırıldığında spor ve doğa temalı paylaşımların yüksek etkileşim üretmesi; buna karşın ürün tanıtımının düşük kalması da, kısa formatta “heyecan/merak/izleme hazzı” sağlayan anlatıların daha fazla paylaşılma, kaydedilme ve yorum alma eğiliminde olduğunu düşündürmektedir. Duygusal tonda ise nötr içeriğin en yüksek etkileşimi sağlaması (yaklaşık %4,79) dikkat çekicidir; aşırı olumlu tonun en düşük kalması (yaklaşık %1,66), izleyicinin aşırı olumlu çerçeveleri yapay veya reklamvari algılayabilmesiyle açıklanabilir. Kuramsal düzeyde çalışma,

Kullanımlar ve Doyumlar Kuramı'nın eğlence/merak/katılım güdülerini, Sosyal Bulunuşluk Kuramı'nın insani ipuçlarının yakınlık ve dikkat üzerindeki etkisini, Parasosyal Etkileşim Kuramı'nın ünlü figürlerin psikolojik yakınlık yaratmasını, Sosyal Kimlik Kuramı'nın kimlik uyumu/aspirasyonel benzeşim süreçlerini ve Duygusal Bulaşma Kuramı'nın duygu aktarımını Reels bağlamında birlikte yorumlayarak bütüncül bir açıklama sunmaktadır. Uygulama açısından ise markalar ve içerik üreticileri için açık öneriler üretmektedir: tanınmış yüzlerin veya güven uyandıran figürlerin kullanımı; insan merkezli, hikâye taşıyan ve dinamik eylem içeren videoların tasarlanması; salt tanıtım yerine bağ kurduracak bağlam ve sosyal ipuçlarının eklenmesi; duygusal tonu aşırı zorlamadan daha doğal ve erişilebilir bir üslup tercih edilmesi etkileşimi artırabilir. Bununla birlikte araştırma, örneklemin üç aylık dönem ve yüksek takipçili küresel hesaplarla sınırlı olması nedeniyle genellenebilirlik açısından kısıt taşımaktadır. Gelecek çalışmaların farklı ülkeler, daha küçük ölçekli içerik üreticileri, daha uzun zaman aralıkları ve TikTok/YouTube Shorts gibi platformlar arası karşılaştırmalarla modeli genişletmesi önerilmektedir. Genel olarak bu çalışma, kısa video ekosisteminde etkileşimi belirleyen faktörleri makine öğrenmesiyle görünür kılarak, hem akademik literatüre hem de veri temelli dijital pazarlama uygulamalarına özgün katkı sunmaktadır.