Early Multimodal Prediction of Cross-Lingual Meme Virality on Reddit: A Time-Window Analysis

Sedat Dogan¹, Nina Dethlefs², Debarati Chakraborty¹

¹School of Computer Science, University of Hull, Hull, UK ²Loughborough University, Loughborough, UK s.dogan-2021@hull.ac.uk

Abstract

Predicting the virality of online content remains challenging, especially for culturally complex, fast-evolving memes. This study investigates the feasibility of early prediction of meme virality using a large-scale, cross-lingual dataset from 25 diverse Reddit communities. We propose a robust, data-driven method to define virality based on a hybrid engagement score, learning a percentile-based threshold from a chronologically held-out training set to prevent data leakage. We evaluated a suite of models, including Logistic Regression, XGBoost, and a Multi-layer Perceptron (MLP), with a comprehensive, multimodal feature set across increasing time windows (30-420 min). Crucially, useful signals emerge quickly: our bestperforming model, XGBoost, achieves a PR-AUC > 0.52 in just 30 minutes. Our analysis reveals a clear "evidentiary transition," in which the importance of the feature dynamically shifts from the static context to the temporal dynamics as a meme gains traction. This work establishes a robust, interpretable, and practical benchmark for early virality prediction in scenarios where full diffusion cascade data is unavailable, contributing a novel cross-lingual dataset and a methodologically sound definition of virality. To our knowledge, this study is the first to combine time series data with static content and network features to predict early meme virality.

Introduction

Online platforms host a torrent of content, where memes often serve as a primary mode of communication, cultural commentary, and the spread of information (or misinformation)(Shifman 2014). Rapid dissemination of memes, characterized by their rapid replication and wide reach, is often described metaphorically as viral, similar to the spread of biological viruses but driven by psychological and emotional mechanisms rather than biological mechanisms(Denisova 2020), (Klastrup 2014). Understanding which memes will achieve widespread popularity, aka "go viral", is therefore a complex but important task. It has significant implications for social media platforms, marketers, and researchers studying the dynamics of information diffusion.

Forecasting virality early in a meme's lifecycle (often within minutes or hours) is particularly challenging. Sparse initial data are a key reason. However, early forecasting

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

offers substantial practical advantages. Early identification can enable proactive content moderation, improve recommendation engines, and yield insight into the initial sparks that ignite widespread diffusion(Han, Lappas, and Sabnis 2020),(Gao et al. 2022). The unique nature of memes, their multimodality, cultural specificity, and rapid evolution add complexity compared to predicting the success of other online content(Barnes et al. 2020),(Ling et al. 2021).

Memes often require culturally specific knowledge to correctly decode humor and meaning, with interpretations that vary significantly between cultural groups(Guseynova et al. 2022). Previous research has explored virality prediction using various features and models, including sophisticated deep learning techniques(Chen et al. 2019a). Despite this, several gaps persist. Many studies focus on single platforms or languages. They adopt varying definitions of virality and often lack a granular analysis of how predictive power develops during those crucial initial hours. Although there are specific studies on meme virality(Barnes et al. 2020),(Ling et al. 2021), the establishment of a robust definition of virality and benchmarks for early prediction in diverse communities remains an open area.

This paper aims to bridge some of these gaps. We investigate the following research questions:

- 1. Can we establish a robust, data-driven definition of virality based on a hybrid engagement score, thus avoiding arbitrary thresholds?
- 2. How accurately can we predict meme virality using combined engagement dynamics, network context, and static content features observed within early time windows (e.g. 30, 60, 120 minutes) after posting?
- 3. How does the predictive performance of interpretable models and deep learning baselines change as the observation window lengthens?
- 4. What is the relative importance of different feature categories for early prediction? How does this importance shift over time?

To frame our investigation, we hypothesize that the signals of meme virality are dynamic and follow a phased progression, an "evidentiary transition" where the nature of predictive evidence shifts from static context to dynamic engagement as a meme's lifecycle unfolds.

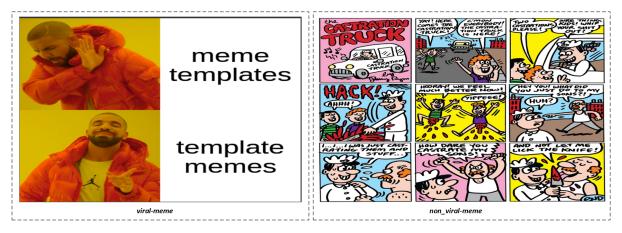


Figure 1: Examples of memes classified as viral (left, Drake meme template) and non-viral (right, multi-panel comic meme) based on our data-driven definition.

To address these questions, we leverage a large-scale, cross-lingual dataset from 25 Reddit communities. We propose and apply a novel method for defining virality based on a hybrid score of weighted engagement and dynamic features. We engineered a comprehensive feature set and systematically evaluated a range of models, including Logistic Regression, XGBoost, and an MLP Neural Network, across multiple time windows (30 to 420 minutes).

Our results show that our feature-rich approach, when paired with a powerful model such as XGBoost, significantly outperforms the deep learning baseline in both accuracy and computational efficiency. Our analysis confirms our hypothesis, revealing a clear temporal shift in the importance of features from the initial context to the observed dynamics of engagement. Figure 1 shows two meme examples that are classified as viral and non_viral by our XGBoost model.

This study makes several key contributions: (1) we introduce a large-scale, cross-lingual dataset for meme analysis; (2) we propose a robust, "data-driven methodology for defining virality" that avoids arbitrary thresholds; (3)we establish strong benchmarks for interpretable prediction, showing that a feature-rich XGBoost model outperforms a deep learning baseline; and (4) we provide novel insights into the timevarying nature of meme success, confirming our "evidentiary transition" hypothesis.

Related Work

Memes serve as cultural capital within communities, strengthening identity through correct interpretation and usage, and these dynamics vary significantly between cultures (Nissenbaum and Shifman 2017). Memes propagate differently across cultures and languages, influenced by local socio-cultural contexts, linguistic nuances, and visual traditions(Ageev, Pushkarev, and Antonenko 2024).

Definitions and Operationalization of Virality

Virality for memes and online content has often been defined through metaphors of rapid and wide dissemination akin to a biological virus, driven by psychological and emotional factors rather than biological ones(Denisova 2020), (Rušinović 2020). Cultural replication perspectives define memes as units of cultural information that spread quickly and broadly online, emphasizing their capacity for rapid replication (Rušinović 2020), (Klastrup 2014). Structural virality distinguishes between content that is spread through large-scale broadcasts and those that is spread through multiple generations of person-to-person sharing(Goel et al. 2015). Furthermore, intertextual and vernacular definitions highlight memes as connective media that link different communities through shared expressive practices and cultural references(Zanette, Blikstein, and Visconti 2019).

Operationalization commonly involves measurable engagement metrics such as shares, likes, comments and views(Barnes et al. 2020), (Smitha, Sendhilkumar, and Mahalakshmi 2018), (Klastrup 2014). Some studies incorporate user behavior intentions to predict virality in conjunction with actual behaviors(Alhabash and McAlister 2015). Furthermore, machine learning approaches that take advantage of structural analyses, particularly those that use network diffusion patterns, offer predictive insights into meme virality (Goel et al. 2015).

Early Prediction of Online Content Virality

Predicting online content virality shortly after publication remains a challenge, but critical for proactive content management. Recent research indicates that early engagement patterns, content characteristics, and network dynamics significantly aid in accurate early stage predictions(Zadeh and Sharda 2022), (Xu and Qian 2023). Temporal models such as multivariate Hawkes processes and epidemic-type diffusion models have been effective in capturing bursty, cascading user engagement, achieving notable predictive accuracy shortly after posting(Zadeh and Sharda 2022), (Xu and Qian 2023).

Deep learning approaches, particularly those integrating temporal, user-centric, and textual characteristics with attention mechanisms, have demonstrated strong predictive performance on platforms such as Twitter and Weibo(Chen et al. 2019a),(Xu and Qian 2023). Node embedding algorithms further enhance predictions by identifying latent node influences, achieving significant early-stage accuracy(Lu and Szymański 2017). Furthermore, cascade graph learning frameworks, such as ViralGCN, effectively integrate spatialtemporal cascade structures for accurate and interpretable predictions, which take advantage of the precise structure of the user-to-user sharing network ((Xu and Qian 2023).Although powerful, these methods require complete knowledge of the diffusion cascade, which is often unavailable in real-time early prediction scenarios or from standard APIs. Our work addresses a complementary and practical problem: predicting virality using only a post's intrinsic content and aggregate engagement dynamics, which are readily accessible without the full diffusion graph. This feature-based approach offers a different advantage: high interpretability for understanding the factors driving virality, as opposed to the often black-box nature of complex GNNs.

Feature Engineering and Predictive Modeling

Advancements in feature extraction for meme virality prediction, leveraging machine learning, computer vision, and NLP, have markedly improved predictive accuracy(Barnes et al. 2020). Static features such as visual clarity, emotional expression, and concise text correlate highly with virality(Ling et al. 2021). Dynamic features such as early diffusion patterns, community spread, and temporal activity further enhance predictive models significantly(Weng, Menczer, and Ahn 2013), (Hui et al. 2018), (Gao et al. 2022). Integrating multimodal features, including visual, textual, temporal, and network data, has enhanced the predictive precision of content virality models(Barnes et al. 2020).

Hierarchical fusion models merging separately extracted visual (e.g. ResNet), textual (e.g. GloVe and LSTM), and attribute-based features significantly outperform singlemodality methods(Wang et al. 2023). Although historical approaches relied on engineered visual features(Lv et al. 2017; Chen et al. 2019b), current work frequently employs large pre-trained vision language models (VLMs) such as CLIP for tasks ranging from harmful content detection to sentiment analysis(Kumar and Nandakumar 2022; Sultan et al. 2024) Complex machine learning models such as XGBoost, LightGBM, and deep neural networks generally achieve higher predictive accuracy (AUC 0.88), though simpler, interpretable models such as logistic regression and random forests still provide valuable insights into feature importance and model behavior(Barnes et al. 2020), (Tsai and Wu 2022), (Kalra et al. 2022).

Virality prediction datasets are inherently imbalanced due to the rarity of viral content, requiring specialized approaches. Techniques such as oversampling (SMOTE), undersampling, ensemble methods, and cost-sensitive learning significantly improve model performance by balancing datasets and reducing false negatives(Kubus 2020), (Van Den Goorbergh et al. 2022), (Kala et al. 2024).

Research Gaps and Our Contribution

This review highlights several gaps our work aims to address. Firstly, integrating insights from large, complex models (like LLMs or VLMs) into practical, efficient prediction systems remains challenging. Secondly, there is a need for more large-scale, cross-lingual benchmarks for meme virality prediction. Third, while temporal dynamics is known to be important, a systematic analysis of how feature importance shifts across early time windows is less common. Finally, many studies adopt simple arbitrarily threshold-based definitions of virality without accounting for the dynamic nature of virality in their experimental design.

Our study makes several key contributions to address these gaps. We introduce a large-scale, cross-lingual dataset with high-resolution temporal data. Crucially, we tackle the definitional challenge by proposing and implementing a novel data-driven methodology to define virality that is both robust and methodologically sound, learning its parameters from a training set held to prevent data leakage. We then systematically evaluate a range of models, from interpretable linear models to strong baselines such as XGBoost and an MLP neural network, using a comprehensive multimodal feature set. Finally, our analysis moves beyond simple prediction to investigate the temporal evolution of predictive signals, confirming our hypothesis of an "evidentiary transition" in the nature of viral signals over a meme's early lifecycle.

Data and Methodology

This section details our data collection, preprocessing, and the rigorous methodological framework used for predictive modeling. We place a strong emphasis on preventing data leakage to ensure the validity of our results.

Dataset Construction and Scope

A dataset that captures the dynamics of cross-lingual memes is the foundation of this research. We collected it via the official Reddit API (PRAW) between March 21 and June 03, 2025. (Figure 2).

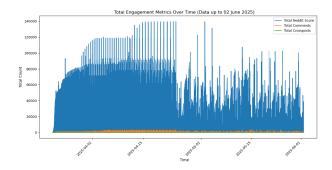


Figure 2: Total engagement metrics of all the collected memes over the tracking time.

For each meme post (j), we gather standard metadata (e.g. title, anonymized author, timestamp, URL of the media) and the subreddit context (subscriber count N_i). A key

element was tracking engagement metrics: score s(t), comments c(t), crossposts x(t). We tracked these with high temporal resolution via dynamic sampling (5-min intervals initially, decreasing frequency later). This strategy helped us track memes life trajectory and spread across the Reddit internal ranking categories, which is crucial for early prediction, while also capturing longer-term trends. The final recorded engagement metrics informed how we defined the target variable.

To ensure diversity, we collected data from 25 distinct meme-centric subreddits across eight language groups. English, German, Turkish, Nordic (Swedish, Danish, Norwegian), French, Spanish (including Latin American variations), Portuguese (including Brazilian Portuguese) and Italian. Detailed information on these subreddits, including their subscriber counts and representation in our dataset, is provided in Table 7 in the appendix. Table 1 shows the distribution of media types between language groups after filtering.

Table 1: Distribution of Media Types Across Language Groups. Note: These counts represent the distribution after the final filtering step (e.g., requiring > 24h tracking).

Language	Image	Video	GIF	Text	Audio	Total
English	27985	2802	755	63	160	31765
German	2345	85	32	47	13	2522
Turkish	610	241	11	118	50	1030
Nordic	618	71	10	183	31	913
French	508	49	9	42	8	616
Spanish	370	183	5	21	16	595
Portuguese	199	58	0	3	1	261
Italian	70	25	1	1	0	97
Total	32705	3514	823	478	279	37799

Standard data cleaning removed moderated posts, posts without a valid URL and posts that did not have at least 24 hours of continuous engagement tracking. Our process adhered to Reddit's Terms of Use, ensured user anonymity, and received faculty approval for responsible conduct. We initially collected 71,040 unique meme posts through the Reddit API. After applying quality filters (removing moderated posts, posts without a media link, and posts with less than 24 hours of continuous engagement tracking), our final dataset contains 37,799 unique meme posts with more than 1 million tracking points.

Target Variable Definition: A Data-Driven Approach

To prevent data leakage, our definition of virality was derived exclusively from the training portion of our dataset and then applied to the held-out test set. Instead of using arbitrary thresholds, we developed a data-driven approach that reflects both the volume and dynamics of engagement. This involved the following steps, performed only on the training data:

1. Normalization by Community Size: We normalized the raw metrics $(k \in \{s, c, x\})$ for each post by its subreddit subscriber count N_i . The 99th percentile cap (P_{99}) used

to mitigate outliers was calculated from the distribution of training data.

$$m_{j,k}(t) = \min\left(\frac{k_j(t)}{N_i} \times 100000, P_{99_train}\left(\frac{k}{N} \times 100000\right)\right)$$
(1)

- 2. **Hybrid Engagement Weighting:** To estimate the relative importance of different engagement signals, we trained an auxiliary Random Forest model on the training set. This model used both normalized volume metrics (score, comments, crossposts) and key dynamic features (e.g., peak velocity, acceleration, time to takeoff) from early time windows to predict a preliminary target (e.g top 5% by unweighted final sum). The importance of the resulting features, averaged between windows, yielded a set of hybrid weights (β_k) for each feature k. This data-driven method confirmed that the normalized score $(\beta_{score} = 1.0)$ and comments $(\beta_{comments} = 0.44)$ were the most important volume signals, while dynamic features such as peak velocity $(\beta_{peak_vel} = 0.14)$ also contributed significantly.
- 3. **Final Hybrid Score:** We calculated a composite Hybrid Score $(HS_{j,final})$ for each post using its final engagement features $(f_{j,k})$ and the derived hybrid weights (β_k) :

$$HS_{j,final} = \sum_{k \in \text{features}} \beta_k \cdot f_{j,k}$$
 (2)

4. **Threshold Identification:** To objectively define our viral class, we applied K-Means clustering (MacQueen 1967) (k=2) to the $(HS_{j,final})$ distribution. The resulting boundary $(\tau \approx 300.27)$ served as our data-driven threshold (Figure 3). Clustering provided a natural separation.

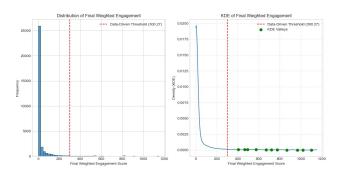


Figure 3: Distribution of Final Hybrid Scores across the training set posts. The vertical line indicates the data-driven virality threshold identified by K-Means.

5. Final Target Variable Assignment: We used this single threshold, τ_{train} , to assign the binary target is_viral to all posts in both the training and test sets.

$$is_viral_j = \begin{cases} 1 & \text{if } HS_{j,final} \ge \tau_{train} \\ 0 & \text{if } HS_{j,final} < \tau_{train} \end{cases}$$
 (3)

This procedure ensures that no information from the test set influences the definition of virality, making our evaluation a true simulation of real-world prediction.

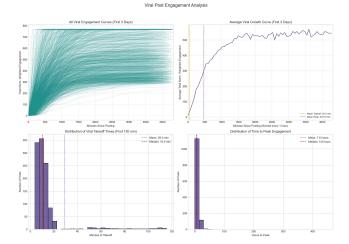


Figure 4: Viral memes lifespan trajectories. Top-left: Individual engagement curves for all viral posts. Top-right: Average viral growth curve. Bottom-left: Distribution of viral take-off times. Bottom-right: Distribution of time to reach peak engagement velocity.

Table 2: Example Features by Modality.

Modality	Example Feature(s)
Temporal	Hour Of Day, Peak Velocity (Window), Burst Count (Window)
Network	Category Transitions, Time To Top, Author Total Karma
Visual	Key Objects, Facial Expression, Template Name, Panels
Textual	Text Sentiment, Title Word Count, Text Language, Text Tone
Contextual	Meme Template, Target Audience, Offensiveness

Exploratory Analysis Findings

Initial exploration of the full dataset confirmed its multimodal nature and revealed interesting patterns. Static images were prevalent, while text-based memes were common in specific language groups (Table 1). as illustrated in the appendix (Figures 7 through 16), characteristic distributions and virality rates vary significantly between language groups and over time, underscoring the need for context-sensitive models. We analyze engagement trajectories for posts ultimately classified as viral, revealing consistent patterns (Figure 4). Viral memes typically showed a rapid take-off (mean: 29 min), but reached maximum engagement velocity much later (mean: 7.6 hours), suggesting a window for early prediction. Our final definition of virality classified $\approx 4.8\%$ of posts as viral, confirming a significant class imbalance that guided our choice of models and metrics.

Feature Engineering

We engineered features that capture dynamic engagement and static characteristics, ensuring that all features for a time window W used only data available up to W. All feature

scaling and normalization parameters were fitted to the training data and then used to transform the test data. The final characteristics are divided into several groups of modality (Table 2).

- **Temporal Dynamics:** Features summarizing participation up to W, including aggregates, counts, and estimates of velocity (v(t)) and acceleration (a(t)). Figure 5 illustrates the average dynamic trajectories.
- Network Context: Static author (karma, age) and characteristics of community structure (e.g., category transitions).
- LLM-Derived Static Features: To extract nuanced semantic information, we used a multimodal LLM (Gemini 2.0 Flash Thinking (Gemini Team and Google 2023)) to generate structured JSON covering visuals (e.g. object detection), text (e.g. OCR) and context (e.g. cultural references).

Crucially, all features were calculated so that for any given post and time window, only data available up to that point from that same post were used. No information from the test set or from future states of any post was used during feature extraction.

Experimental Setup

This section details our experimental setup, including our rigorous data-splitting protocol, the models evaluated, and the metrics used.

Data Splitting and Experimental Integrity: To ensure that our models are evaluated on truly unseen data, we split this filtered dataset chronologically. All posts from March 21 to May 15, 2025 (approximately 80% of the data) were used for the training set, and posts from May 16 to June 3 (approximately 20%) formed the held-out test set (Table 3). This chronological split prevents data leakage and ensures a realistic evaluation scenario in which models are tested on future data relative to their training period.

Table 3: Chronological Split of the Final Filtered Dataset

Split	Date Range	Count	Percent
Training Set Test Set	Mar 21–May 15, 2025 May 16–Jun 3, 2025	30,239 7,560	80.0% 20.0%
Total	Mar 21–Jun 3, 2025	37,799	100%

Preprocessing Pipeline: For each time window, we constructed a feature set that combines static attributes with the dynamic features available up to that point. We used a robust preprocessing pipeline where numeric features were imputed with the median and standardized, while categorical features were imputed with a 'missing' value, and then one hot encoded. All imputation and scaling parameters were learned from the training data only.

Handling Class Imbalance: Given the significant class imbalance inherent in virality prediction (\approx 4.8% viral posts), we address this challenge by using class weights in our models to increase the penalty for misclassifying the minority (viral) class.

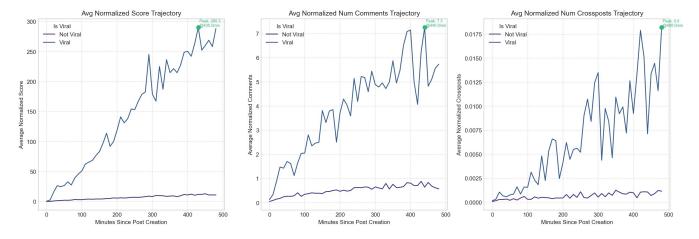


Figure 5: Average normalized engagement trajectories (Score, Comments, Crossposts) for posts classified as Viral vs. Non-Viral over the first 500 minutes. Viral posts show distinctly higher and faster-rising engagement across all metrics.

Models Evaluated: We evaluated three machine learning models: Logistic regression (LR) as a strong linear baseline, XGBoost as a state-of-the-art tree-based model, and a multilayer perceptron (MLP) neural network as a deep learning baseline. The hyperparameter configurations for each model are detailed in Table 4.

- Logistic Regression (LR): A linear model chosen for its high interpretability.
- XGBoost: A powerful and highly efficient gradient boost implementation, serving as a strong tree-based baseline.
- MLP Neural Network: A Multi-layer Perceptron with two hidden layers, serving as a standard deep learning baseline.

Table 4: Model configurations and hyperparameters

Model	Hyperparameter Configuration penalty='12' C=1.0 solver='liblinear' class_weight='balanced' random_state=42	
Logistic Regression		
XGBoost	scale_pos_weight=#neg/#pos eval_metric='logloss' use_label_encoder=False random_state=42 n_jobs=-1	
MLP Neural Network	hidden_layer_sizes=(100, 50) activation='relu' solver='adam' max_iter=500 early_stopping=True random_state=42	

Evaluation Metrics

To obtain a stable performance estimate, we report the results of a five-fold stratified cross-validation performed on

the training set. Our primary metric is the area under the precision recall curve (PR AUC), which is well suited for unbalanced classification tasks (Davis and Goadrich 2006). We also report the ROC AUC and the F1 score for a complete assessment. The final best-performing model for each time window is then evaluated in the held-out test set to report the final generalization performance.

Experiments and Results

This section details the performance of our predictive models, reports the findings of our feature ablation studies, and analyzes the evolution of feature importance over time. All results reported here were generated using the strict split test train and the cross-validation procedure defined in the earlier steps to ensure methodological rigor.

Main Results: Performance Across Time Windows

Table 5 summarizes the performance of our three models (LR, XGBoost, and the MLP Neural Network baseline) in increasing observation windows. The results demonstrate a clear and consistent trend: as more engagement data becomes available over time, the predictive power of all models increases significantly.

XGBoost emerges as the strongest performer in all time windows and metrics, highlighting the effectiveness of gradient boosting for this task. It achieves a PR-AUC of 0.52 in just 30 minutes, rising to a robust 0.82 after 420 minutes. The MLP neural network neural network also consistently outperforms the simpler Logistic Regression baseline, indicating that nonlinear relationships are crucial to accurately predict virality. Furthermore, the duration column highlights the computational efficiency of XGBoost, which consistently provides the best performance in the shortest amount of time.

Ablation Study: Importance of Feature Modalities

To isolate the contribution of different categories of features, we performed an ablation study on our best-performing

Table 5: Test Set Performance of LR, XGBoost, and MLP Neural NetworkBaseline Across Time Windows. The table highlights the trade-off between predictive power (PR AUC) and computational cost (Duration). Best performance for each metric is in bold.

Time Window (min)	Model Type	PR AUC	ROC AUC	F1 Score	Duration (s)
30	Logistic Regression	0.35	0.85	0.36	16.95
30	MLP Neural Network	0.41	0.86	0.24	1121.21
30	XGBoost	0.52	0.93	0.46	7.05
60	Logistic Regression	0.44	0.86	0.39	17.38
60	MLP Neural Network	0.47	0.87	0.35	1142.22
60	XGBoost	0.58	0.94	0.49	7.04
120	Logistic Regression	0.54	0.88	0.47	16.95
120	MLP Neural Network	0.57	0.89	0.51	1124.60
120	XGBoost	0.65	0.95	0.54	7.47
180	Logistic Regression	0.61	0.90	0.52	13.78
180	MLP Neural Network	0.63	0.91	0.54	1385.64
180	XGBoost	0.70	0.95	0.58	7.19
240	Logistic Regression	0.66	0.91	0.57	14.35
240	MLP Neural Network	0.67	0.92	0.59	1459.23
240	XGBoost	0.75	0.96	0.63	7.21
300	Logistic Regression	0.70	0.92	0.60	17.44
300	MLP Neural Network	0.71	0.932	0.63	1132.59
300	XGBoost	0.78	0.96	0.65	7.35
360	Logistic Regression	0.74	0.93	0.61	17.58
360	MLP Neural Network	0.73	0.93	0.66	1139.11
360	XGBoost	0.80	0.97	0.68	7.65
420	Logistic Regression	0.76	0.94	0.62	19.40
420	MLP Neural Network	0.76	0.94	0.68	1097.85
420	XGBoost	0.82	0.97	0.69	7.55

Table 6: XGBoost Ablation Study Test Set Results (120-Min Window).

Ablation Scenario (XGBoost)	PR AUC	ROC AUC
Baseline (All Features)	0.65	0.951
Exclude Contextual (LLM)	0.66	0.95
Exclude Visual (LLM)	0.65	0.94
Exclude Textual (LLM)	0.63	0.94
Exclude Network	0.56	0.90
Exclude Temporal	0.43	0.92

model, XGBoost. Although we evaluated a MLP neural network baseline, its consistently lower performance across all time windows (see Table 5) makes interpreting its importance of features less reliable to establish a general hierarchy of predictive signals. Therefore, we focus the ablation analysis on the XGBoost model to provide the clearest insight into the factors driving the successful prediction.

For the 120-minute window, we systematically trained the model, while excluding one feature modality at a time. Table 6 shows the results. The findings confirm the critical importance of temporal and network characteristics. Excluding temporal features caused the largest drop in PR-AUC (from 0.65 to 0.43), demonstrating that the early dynamic

trajectory of a meme is the single most important predictor. Removing network features also significantly degraded performance (PR-AUC drop to 0.56), emphasizing the predictive power of the initial author and community context. In contrast, removing the LLM-derived static features (Visual, Textual, Contextual) had a much smaller impact, suggesting that they provide valuable but less critical context compared to the dynamic and network signals.

Feature Importance Analysis Over Time

We further examine how feature importance evolves using feature importance from the XGBoost models across time windows. Figure 6 reveals a more nuanced temporal pattern than a simple linear shift. In the earliest windows (30-120 min), static features related to the post's content and context, specifically 'Textual' and 'Network' features, are highly prominent. This suggests that initial predictions rely heavily on the meme's textual content and the author's standing.

As more engagement data accumulate in the middle windows (180-300 min), the 'Temporal' features, which capture the dynamics of user interaction like velocity and acceleration, surge in importance, becoming the dominant predictive modality. This indicates a transition from predicting based on what the content is to predicting based on how it is behav-

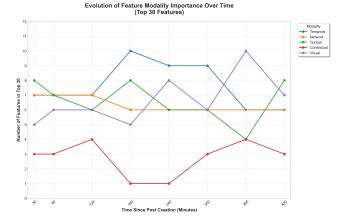


Figure 6: Evolution of feature modality importance (count in Top 30 features) for XGBoost models trained at different time windows.

ing. In the later stages (360-420 min), we observe a resurgence in the importance of 'Visual' and 'Textual' features, suggesting that for a meme to have long-term viral potential, the intrinsic quality of its visual and textual content becomes critical again, complementing the established dynamic trajectory.

Conclusion and Discussion

Our results confirm that the early prediction of meme virality is not only feasible but can be achieved with a high degree of accuracy. Using a robust methodologically robust leak-free pipeline, our analysis shows that performance improves substantially as more engagement data accumulate (Table 5). Our strongest model, XGBoost, consistently outperforms both a standard deep learning baseline (MLP) and a simpler linear model, suggesting that gradient-boosted trees are particularly well-suited for this feature-rich prediction task. Meaningful predictions (PR AUC > 0.52) are possible in 30 minutes, becoming highly reliable (PR AUC > 0.82) by the 7-hour mark. Although PR AUC is our main metric due to class imbalance, ROC AUC scores achieved 0.92 even without temporal features (Table 6) compare favorably with previous work focusing mainly on static features (e.g. 0.68-0.87 (Barnes et al. 2020), (Ling et al. 2021). Of course, direct comparison is limited by differing datasets and methodolo-

A key contribution of this work is insight into the timevarying nature of predictive signals. Our feature importance analysis (Figure 6) reveals a clear "evidentiary transition" in the way virality manifests. We observe a phased progression:

- 1. **The Seeding Phase (0-120 mins):** Early predictability is dominated by static context, namely *Network* and *Textual* features. This highlights the initial importance of the author's reputation and the meme's textual framing.
- 2. **The Ignition Phase (180-300 mins):** As engagement data cumuluate, a distinct shift occurs and dynamic *Temporal* features become paramount. This reflects the establishment of a positive feedback loop where the meme's

- observed performance is the strongest predictor.
- 3. **The Sustain Phase (360+ mins):** In later stages, the intrinsic quality of the content reemerges as critical, with *Visual* and *Textual* features regaining prominence to maintain long-term momentum.

This dynamic interplay is further supported by our ablation study (Table 6), which confirms that temporal and network characteristics are the most critical components for early prediction.

In conclusion, this study makes several key contributions: (1) we introduce a large-scale, cross-lingual dataset for meme analysis; (2) we propose a robust, "data-driven methodology for defining virality" that avoids arbitrary thresholds; (3) we establish strong benchmarks for interpretable prediction, showing that a feature-rich XGBoost model outperforms a deep learning baseline in both performance and efficiency; and (4) we provide novel insights into the time-varying nature of meme success, confirming our "evidentiary transition" hypothesis. It is also important to note that our approach is tailored to predicting virality from a post's intrinsic features and temporal engagement, a common scenario where the user-to-user diffusion graphs required by models like ViralGCN are unavailable, thus justifying our feature-based methodology.

Limitations

The findings of this study are primarily limited by the scope of the dataset; Reddit - during March-June 2025. Validation across platforms and timeframes is certainly needed. In addition, while diverse, the data set exhibits imbalances. Fewer examples exist for some languages (e.g., Italian, Portuguese) and media types (e.g., Audio, Text; see Table 1). This potentially limits the generalizability of findings specific to these minority categories. We used an advanced LLM as a feature extraction tool based on its known capabilities.

However, this study did not perform a direct performance comparison against baselines of non-LLM static characteristics. Quantifying the specific contribution of these LLMderived features relative to simpler alternatives is crucial future work. Reliance on a specific LLM also brings considerations of availability, cost (though there is a free tier (Gemini Team and Google 2023)) and potential biases. Methodological limitations include the potential Gini importance bias and the specific operationalization of virality we chose. Ethically, the potential for misuse of virality prediction (for example, for manipulation or amplifying harmful content) warrants careful consideration. Potential biases encoded within the data or models also need attention. Methodologically, while we benchmark against strong baselines (XGBoost, MLP), we did not implement more complex, graph-based architectures like GNNs, as our dataset is not structured as a diffusion cascade.

Future work should focus on refining LLM integration (e.g., prompt optimization) and exploring advanced model architectures (perhaps time-sensitive attention or graph models). A key priority remains to rigorously compare LLM-derived features against computationally cheaper alternatives (such as CLIP or standard OCR) to establish their

cost-benefit relationship. Targeted data collection for underrepresented languages and media types could enable a more robust analysis of these specific categories.

Crucially, future research should explicitly address fairness and bias mitigation in prediction models. Considering the broader social impacts and ethical guidelines for the deployment of such systems is also essential.

Our work focuses primarily on predictive accuracy and does not explicitly model the real-world cost of misclassification. The implications of a false positive and a false negative can be vastly different and are highly dependent on the application scenario. For example:

In a content moderation context, a false negative (failing to identify a truly viral piece of harmful misinformation) could have severe societal consequences by allowing it to spread unchecked. In this case, high recall is paramount.

Conversely, in a recommendation system, a false positive (promoting a non-viral meme to a large audience) might incur an opportunity cost and degrade user experience. Here, high precision would be preferred.

Our current model, with its fixed decision threshold, does not optimize for these asymmetric costs. A valuable direction for future work is to move beyond a single accuracy metric and incorporate cost-sensitive learning techniques, allowing the decision threshold to be tuned for specific downstream applications and their unique ethical and practical requirements.

Data and Code Availability

The dataset, the code, and the specific LLM prompt used to extract the characteristics are available at https://github.com/sdogan13/Meme-Analysis-and-Virality-Prediction.

References

- Ageev, S.; Pushkarev, E.; and Antonenko, N. 2024. Cognitive underpinnings of misperceptions in morphed humor. *Russian Journal of Linguistics*.
- Alhabash, S.; and McAlister, A. 2015. Redefining virality in less broad strokes: Predicting viral behavioral intentions from motivations and uses of Facebook and Twitter. *New Media & Society*, 17: 1317–1339.
- Barnes, K.; Riesenmy, T.; Trinh, M. D.; Lleshi, E.; Balogh, N.; and Molontay, R. 2020. Dank or not? Analyzing and predicting the popularity of memes on Reddit. *Applied Network Science*, 6.
- Chen, G.; Kong, Q.; Xu, N.; and Mao, W. 2019a. NPP: A neural popularity prediction model for social media content. *Neurocomputing*, 333: 221–230.
- Chen, J.; Liang, D.; Zhu, Z.; Zhou, X.; Ye, Z.; and Mo, X. 2019b. Social Media Popularity Prediction Based on Visual-Textual Features with XGBoost. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19)*, 2322–2330.
- Davis, J.; and Goadrich, M. 2006. The Relationship Between Precision-Recall and ROC Curves. In *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, 233–240. ACM.

- Denisova, A. 2020. How to Define 'Viral' for Media Studies? Westminster Papers in Communication and Culture.
- Gao, X.; Jia, X.; Yang, C.; and Chen, G. 2022. Using Survival Theory in Early Pattern Detection for Viral Cascades. *IEEE Transactions on Knowledge and Data Engineering*, 34: 2497–2511.
- Gemini Team; and Google. 2023. Gemini: A Family of Highly Capable Multimodal Models. arXiv preprint arXiv:2312.11805. Accessed: May 1, 2025.
- Goel, S.; Anderson, A.; Hofman, J.; and Watts, D. 2015. The Structural Virality of Online Diffusion. *Manag. Sci.*, 62: 180–196.
- Guseynova, L.; Dugalich, N.; Lomakina, O.; Neliubova, N. Y.; and Ebzeeva, Y. 2022. The Reflection of the Socio-Cultural Context in Russian, French and Azerbaijani Internet Memes. *RUDN Journal of Language Studies, Semiotics and Semantics*.
- Han, Y.; Lappas, T.; and Sabnis, G. 2020. The Importance of Interactions Between Content Characteristics and Creator Characteristics for Studying Virality in Social Media. *Inf. Syst. Res.*, 31: 576–588.
- Hui, P.-M.; Weng, L.; Shirazi, A. S.; Ahn, Y.-Y.; and Menczer, F. 2018. Scalable Detection of Viral Memes from Diffusion Patterns. 197–211.
- Kala, M.; Jain, S.; Singh, A.; and Krishnan, N. 2024. Addressing class imbalance in avalanche forecasting. *Cold Regions Science and Technology*.
- Kalra, S.; Kumar, C. H. S.; Sharma, Y.; and Chauhan, G. 2022. Comparative Analysis of Various Machine Learning Based Techniques for Predicting the Virality of Tweets. 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 601–605.
- Klastrup, L. 2014. "HEARTWARMING MOMENT OF THE DAY": GLOBAL SHARING OF VIRAL CONTENT ON FACEBOOK. 4.
- Kubus, M. 2020. Evaluation of Resampling Methods in the Class Unbalance Problem. *Econometrics*, 24: 39–50.
- Kumar, G.; and Nandakumar, K. 2022. Hate-CLIPper: Multimodal Hateful Meme Classification based on Crossmodal Interaction of CLIP Features. arXiv preprint arXiv:2210.05916.
- Ling, C.; AbuHilal, I.; Blackburn, J.; De Cristofaro, E.; Zannettou, S.; and Stringhini, G. 2021. Dissecting the Meme Magic: Understanding Indicators of Virality in Image Memes. *Proceedings of the ACM on Human-Computer Interaction*, 5: 1–24.
- Lu, X.; and Szymański, B. 2017. Predicting Viral News Events in Online Media. 2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), 1447–1456.
- Lv, J.; Liu, W.; Zhang, M.; Gong, H.; Wu, B.; and Zheng-Jun Zha. 2017. Multi-feature Fusion for Predicting Social Media Popularity. In *Proceedings of the 25th ACM international conference on Multimedia (MM '17)*, 1725–1732.
- MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of*

the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, 281–297. University of California Press.

Nissenbaum, A.; and Shifman, L. 2017. Internet memes as contested cultural capital: The case of 4chan's /b/ board. *New Media & Society*, 19: 483–501.

Rušinović, D. 2020. Internetski memi u sociologijskoj perspektivi.

Shifman, L. 2014. The Cultural Logic of Photo-Based Meme Genres. *Journal of Visual Culture*, 13: 340–358.

Smitha, E.; Sendhilkumar, S.; and Mahalakshmi, G. 2018. A Methodology for Meme Virality Prediction. *International Conference on Computer Networks and Communication Technologies*.

Sultan, T.; M. M. R. Rony; Md Rifatul Islam; Aldosary, S.; and Walid El-Shafai. 2024. MemesViTa: A Novel Multimodal Fusion Technique for Troll Memes Identification. *IEEE Access*, 12: 177811–177828.

Tsai, M.-J.; and Wu, Y. 2022. Predicting online news popularity based on machine learning. *Comput. Electr. Eng.*, 102: 108198.

Van Den Goorbergh, R.; Smeden, M.; Timmerman, D.; and Calster, B. 2022. The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression. *Journal of the American Medical Informatics Association: JAMIA*, 29: 1525–1534.

Wang, J.; Yang, S.-X.; Zhao, H.; and Yang, Y. 2023. Social media popularity prediction with multimodal hierarchical fusion model. *Comput. Speech Lang.*, 80: 101490.

Weng, L.; Menczer, F.; and Ahn, Y.-Y. 2013. Virality Prediction and Community Structure in Social Networks. *Scientific Reports*, 3.

Xu, Z.; and Qian, M. 2023. Predicting Popularity of Viral Content in Social Media through a Temporal-Spatial Cascade Convolutional Learning Framework. *Mathematics*.

Zadeh, A.; and Sharda, R. 2022. How Can Our Tweets Go Viral? Point-Process Modelling of Brand Content. *Inf. Manag.*, 59: 103594.

Zanette, M.; Blikstein, I.; and Visconti, L. 2019. INTER-TEXTUAL VIRALITY AND VERNACULAR REPER-TOIRES: INTERNET MEMES AS OBJECTS CONNECT-ING DIFFERENT ONLINE WORLDS. Revista de Administração de Empresas.

Appendix

This section details the subreddit information, static and dynamic features used in the modality-specific analysis, grouped by their conceptual modality. Visual, contextual, and textual characteristics were derived using the LLM prompting methodology described in Section and summarized in Appendix 14.

Subreddit Distributiton

Table 7: Detailed Subreddit Distribution by post count

Subreddit	Post Count	Percentage (%)	Subscribers
memes	8,745	23.41	35,495,085
formuladank	6,075	16.26	1,004,277
shitposting	4,463	11.95	2,623,506
antimeme	3,677	9.84	1,031,079
historymemes	2,174	5.82	12,533,773
me_irl	2,061	5.52	7,709,088
dankmemes	1,927	5.16	5,876,517
lotrmemes	1,090	2.92	2,975,156
bonehurtingjuice	756	2.02	960,919
bikinibottomtwitter	400	1.07	4,724,787
surrealmemes	211	0.56	905,018
raimimemes	173	0.46	474,809
wholesomemes	4	0.01	699,206
wholesomemenes	1	0.0	18,334,345
rance	524	1.4	705,861
moi_dlvv	3	0.01	49,257
ich_iel	2,652	7.1	2,017,877
okbrudimongo	602	1.61	102,674
deutschememes	265	0.71	263,213
unket	691	1.85	111,879
dankmark	118	0.32	82,635
eu_nvr	105	0.28	882,683
dankgentina	483	1.29	184,365
yo_elvr	1	0.0	23,212
burdurland	153	0.41	503,976

Feature Definitions by Modality

Exploratory Analysis of Virality Rates

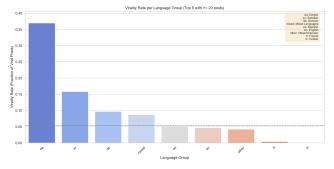


Figure 7: Virality Rate per Language Group.

Table 8: Contextual Features (LLM-Derived).

Contextual Features (LLM-Derived)				
Feature Name	Data Type			
Offensiveness (Is Offensive,	bool, object			
Type)				
Cultural Reference Type	object			
Primary Topic	object			
Target Audience	object			
Meme Type	object			
Analyzed Media Type	object			
Title Media Coherence	object			
Controversy (Score, Type)	int64, object			
Emotional Resonance	object			
Humor Type	object			
Insight Commentary Score	int64			
Novelty Uniqueness Score	int64			
Profanity Level	object			
Relatability Score	int64			
Format/Presentation (Effort,	object, int64			
Simplicity, Appeal, Clarity)				
Social Context (Platform,	object			
Shareability, Currency,				
Trend)				

Table 9: Network Features.

Network Features	
Feature Name	Data Type
Author Account Age Days	float64
Author Is Premium	int64
Author Karma Per Day	float64
Author Total Karma	int64
Category Sequence Full	object
Category Stability Full	float64
Category Transitions Full	int64
Pct Time In New Full	float64
Progression Pattern Full	object
Promotion Demotion Ratio	float64
Full	
Time To Hot Full	float64
Time To Rising Full	float64
Time To Top Full	float64
Unique Categories Full	int64

Table 10: Textual Features (LLM-Derived).

Textual Features (LLM-Derived)				
Feature Name	Data Type			
Text Language	object			
Text Sentiment Overall	object			
Text Word Count	float64			
Text Image Alignment	object			
Text Tone	object			
Is Title Present	bool			
Title Word Count	int64			
Title Sentiment	object			

Table 11: Visual Features (LLM-Derived).

Visual Features (LLM-Derive	,
Feature Group / Name	Data Type
Media Type	object
Image Dimensions (Height,	float64
Width)	
Key Objects Primary	object
Composition	object
Panels	object
Template Is Variant	bool
Template Name	object
Facial Expression Is Face	bool
Facial Expression Primary	object
Emotion	
Identified Person Is Celebrity	bool
Identified Person Is Character	bool
Identified Character Name	object
Identified Person Celebrity	object
Name	-

Table 12: Temporal Features (Submission Time & Dynamics).

Temporal Features (Submission Time & Dynamic Feature Name	Data Type
Day Of Week	int64
Hour Of Day	int64
Is Weekend	int64
Burst Count (Window)	float64
Engagement AUC (Window)	float64
Min Acceleration (Window)	float64
Momentum Ratio (Window)	float64
Norm. Num Comments (Window)	float64
Norm. Score (Window)	float64
Pct Time In Hot (Window)	float64
Pct Time In New (Window)	float64
Pct Time In Rising (Window)	float64
Pct Time In Top (Window)	float64
Peak Acceleration (Window)	float64
Peak Velocity (Window)	float64
Takeoff Velocity (Window)	float64
Category Snapshot (Window)	object
First Comm Min (Window)	float64
First Cross Min (Window)	float64
First Vote Min (Window)	float64
Half Life Minutes (Window)	float64
Minutes (Window Duration)	float64
Norm. Num Crossposts (Window)	float64
Slope 10min (Window)	float64
Slope 5min (Window)	float64
Time To Peak (Window)	float64
Time To Takeoff (Window)	float64
Timing Entropy (Window)	float64
Transitions Within (Window)	int64
Time In Hot (Window)	float64
Time In New (Window)	float64
Time In Rising (Window)	float64
Time In Top (Window)	float64
Upvote Ratio (Window)	float64

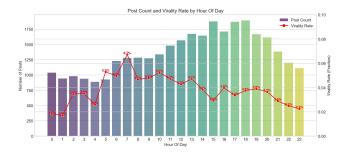


Figure 8: Post Count and Virality Rate by Hour Of Day (UTC).



Figure 9: Post Count and Virality Rate by Day Of Week.

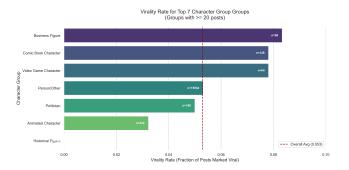


Figure 10: Virality Rate by Character Group.

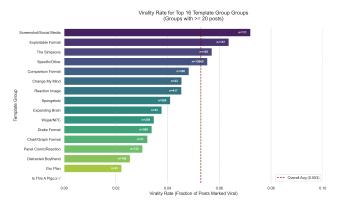


Figure 11: Virality Rate by Template Group.

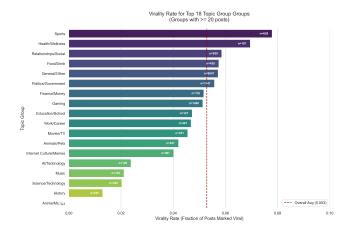


Figure 12: Virality Rate by Topic Group.

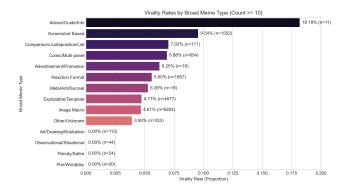


Figure 13: Virality Rate by Broad Meme (Count >= 10).

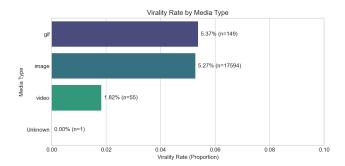


Figure 14: Virality Rate by Media Type.

LLM Prompt Summary for Feature Extraction

The following summarizes the core instructions provided to the Gemini 2.0 Flash Thinking model via the API to extract structured visual, contextual, and textual characteristics, as described in Section. The complete prompt, including the detailed JSON structure and specific categorical/ordinal options for the 'virality_factors', is available in the project's GitHub repository (see the Data and Code Availability section). System Role: Professional meme analyst specializing in identifying factors contributing to viral potential on platforms such as Reddit. Task: Analyze provided media file and title to extract descriptive characteristics and evaluate potential virality drivers using only the specified categorical / ordinal options. Analysis

Steps:

- Identify key visual elements (people, objects, symbols).
- Note the hierarchy and placement of the text.
- · Recognize known meme templates.
- · For video/GIF, extract key frames.
- For audio, evaluate the tone.
- Examine textual and visual references for cultural cues.
- Describe the layout and color composition.
- Analyze the language and references for the target audience.
- Evaluate any offensive content.
- Assess the severity and category of any offensive content.
- Check the coherence between the post title and the media.
- Fill in the 'virality factors' section by selecting the single best fit from the provided options for each factor (using 'unknown' or 'none' if applicable).

SHAP Feature Importance Analysis

To complement the Gini importance analysis and provide further information on the contributions of characteristics, we conducted a SHAP analysis (SHapley Additive Explanations) for the Random Forest model. SHAP values quantify the marginal contribution of each feature to the prediction for each instance. Figure 15 shows the detailed summary graph of the beeswarm for the full baseline model (420 minute window), illustrating both the magnitude and direction of impact of the top features in all the test

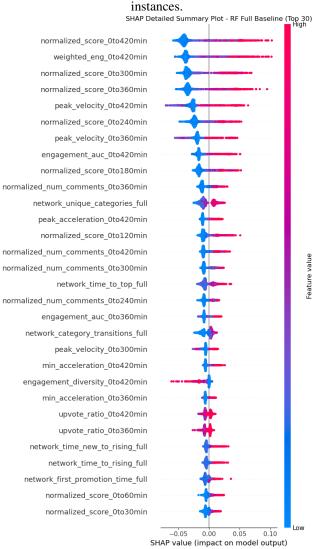


Figure 15: SHAP Detailed Summary Plot (Beeswarm) for RF Full Baseline Model (Top 30 Features).

Feature Distribution of Viral Memes by Language

Cross-Cultural Virality Indicators

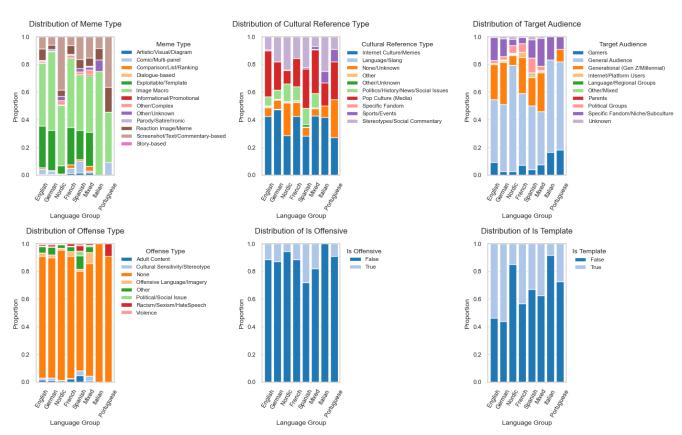


Figure 16: Distribution of selected static features (Meme Type, Cultural Reference Type, Target Audience, Offense Type, Is Offensive, Is Template) across different language groups, highlighting cross-cultural variations.