Mind, Eye, Insight:
Decoding Engagement Patterns in the Self-Improvement YouTube Sphere

Huy Pham
Department of Statistics and Probability, Michigan State University
phamdinh@msu.edu — Website

April 2025

Contents

1 Problem Statement & Motivation **2**
1.1 Problem Statement . . . 2
1.2 Motivation . . . 2
1.3 Gaps in Prior Work . . . 2

2 Dataset **3**
2.1 Collection Protocol . . . 3
2.2 Quality Control and Pre-processing . . . 3

3 Feature Engineering **4**
3.1 Textual Features . . . 4
3.2 Visual Features (LLM-derived) . . . 5
3.3 Temporal & Metadata Features . . . 5

4 Methodology **6**
4.1 Modeling Pipeline . . . 6

5 Evaluation Framework **6**

6 Results **7**
6.1 Exploratory Insights . . . 7
6.2 Model Performance . . . 12
6.2.1 Predicting Engagement Rate . . . 14
6.3 Feature Importance . . . 16
6.3.1 Importance for Engagement Rate (LightGBM) . . . 16

7 Discussion **17**
7.1 Strengths . . . 17
7.2 Limitations . . . 18

8 Conclusion & Future Work **18**

Abstract

This project investigates which multimodal attributes spanning metadata, language, upload timing, and thumbnail design drive audience engagement for English language self-improvement videos on YouTube. A pipeline collected metadata, transcripts, and thumbnails for **7,117** videos returned by 20 carefully-curated queries (e.g. "how to break bad habits", "unlocking potential"). Feature engineering produced textual sentiment and topic mixtures, rich thumbnail descriptors via Gemini 2 Flash, and temporal / channel metrics. Feature engineering produced textual sentiment and topic mixtures, rich thumbnail descriptors via Google's Gemini model (specifically 'gemini-2.0-flash'), and temporal/channel metrics. Ensemble regressors (XGBoost) and tuned dense neural networks achieved an R^2 of approximately **0.67** on held-out test data for predicting `log_view_count`. Exploratory data analysis confirms the positive impact of including faces (especially smiling), moderate text prominence, and high-contrast color palettes in thumbnails, while video age remains the single strongest negative predictor of recent views. The study demonstrates the value of LLM-derived visual features for engagement modeling, highlights current limitations (transcript availability, LLM consistency), and suggests avenues for future work including causal inference and multilingual analysis.

1 Problem Statement & Motivation

1.1 Problem Statement

The digital landscape for self-improvement content is vast and highly competitive. Content creators, educators, and platform designers require evidence-based guidance on how various aspects of a video, from its title and thumbnail to its content and upload schedule, influence viewer behaviour and engagement. This project addresses the core question: *Which specific, measurable content and presentation attributes best predict audience engagement metrics (specifically views, likes, and comments, combined into derived rates) for English-language self-improvement videos on YouTube?*

1.2 Motivation

YouTube serves as a primary platform for individuals seeking self-help and personal development information, accumulating billions of views annually in this category. Understanding the drivers of engagement offers significant benefits:

- **Content Creators:** Can optimize their production and promotion strategies to maximize reach, impact, and channel growth within the self-improvement niche.
- **Platform Designers:** Can refine recommendation algorithms to surface high-quality, engaging, and potentially behaviour-changing content more effectively.
- **Educators & Communicators:** Can leverage insights to design more persuasive and effective digital campaigns for public health or educational purposes.

By identifying key engagement factors, this research aims to provide actionable insights for stakeholders in the digital self-improvement ecosystem.

1.3 Gaps in Prior Work

While numerous studies have analyzed YouTube engagement, many suffer from limitations this project seeks to address:

- **Generality:** Often treat YouTube as monolithic, overlooking niche-specific dynamics like those in self-improvement.
- **Modality Focus:** Tend to concentrate on a single data type (e.g., text only, metadata only), missing crucial cross-modal interactions.
- **Interpretability:** Frequently employ complex "black-box" deep learning models that hinder the extraction of clear, actionable advice for creators.
- **Novel Features:** Few studies have incorporated structured visual features extracted from thumbnails using modern Large Language Models (LLMs).

This project introduces a novel multimodal approach, specifically using LLMs for thumbnail analysis within the self-improvement context, aiming for both predictive accuracy and interpretable insights.

2 Dataset

2.1 Collection Protocol

A custom data collection pipeline was developed using the YouTube Data API v3 and related libraries (‘youtube-transcript-api‘, ‘yt-dlp‘). Using the YouTube Data API v3 we issued 20 keyword queries covering sleep, habit-formation, focus, stress, productivity, and personal growth. The final corpus contains **7,117** unique videos (**3,925** channels).

- **Metadata**: Core information including video title, description, duration (ISO 8601 format), publication timestamp, channel ID, and channel title was collected via the API. Additional metadata like detailed category tags was fetched using ‘yt-dlp‘.
- **Engagement Metrics**: Snapshot data for view counts, like counts, and comment counts were retrieved via the API (as of late March 2025).
- **Transcripts**: English transcripts were automatically fetched using ‘youtube-transcript-api‘. Coverage was approximately **65%** across the dataset, with missing transcripts primarily due to unavailability or non-English content.
- **Thumbnails**: The standard resolution thumbnail URL was collected for each video via the API. The images themselves were downloaded for subsequent visual analysis.

Robust caching mechanisms and exponential back-off strategies were implemented to manage API quotas efficiently, significantly reducing redundant calls (by **82%**).

2.2 Quality Control and Pre-processing

1. **Missing transcripts**: videos without captions were retained; text-based features were imputed with neutral values (zero or `N/A`) so the model can still use the “missingness” signal.

2. **Engagement normalisation**: rows beyond the 98th percentile for `view_count`, `like_count`, `comment_count`, or the initial engagement rate were dropped. A $\log(1+x)$ transform was then applied and the rates recomputed:

$$\text{like\_rate\_log} = \frac{\log(1+\text{like})}{\log(1+\text{views})}, \quad \text{comment\_rate\_log} = \frac{\log(1+\text{comments})}{\log(1+\text{views})},$$

$$\text{engagement\_rate\_log} = 0.2\,\text{like\_rate\_log} + 0.8\,\text{comment\_rate\_log},$$

defaulting to 0 when the denominator is 0.

3. **Missing-value imputation**: text_readability and text_prominence → 0; perceived_emotion and background_complexity → N/A.

4. **Type conversion**: all binary text flags (person_present, is_smiling, etc.) were cast to {0,1}.

5. **Scaling & encoding**: numerical features were standardised; categoricals received one-hot encoding; ordinal variables (e.g. text_readability) kept their intrinsic order.

6. **Data split**: an 85/15 random train–test split was used because the collection window spans only six months; a chronological split is reserved for future, longer datasets.

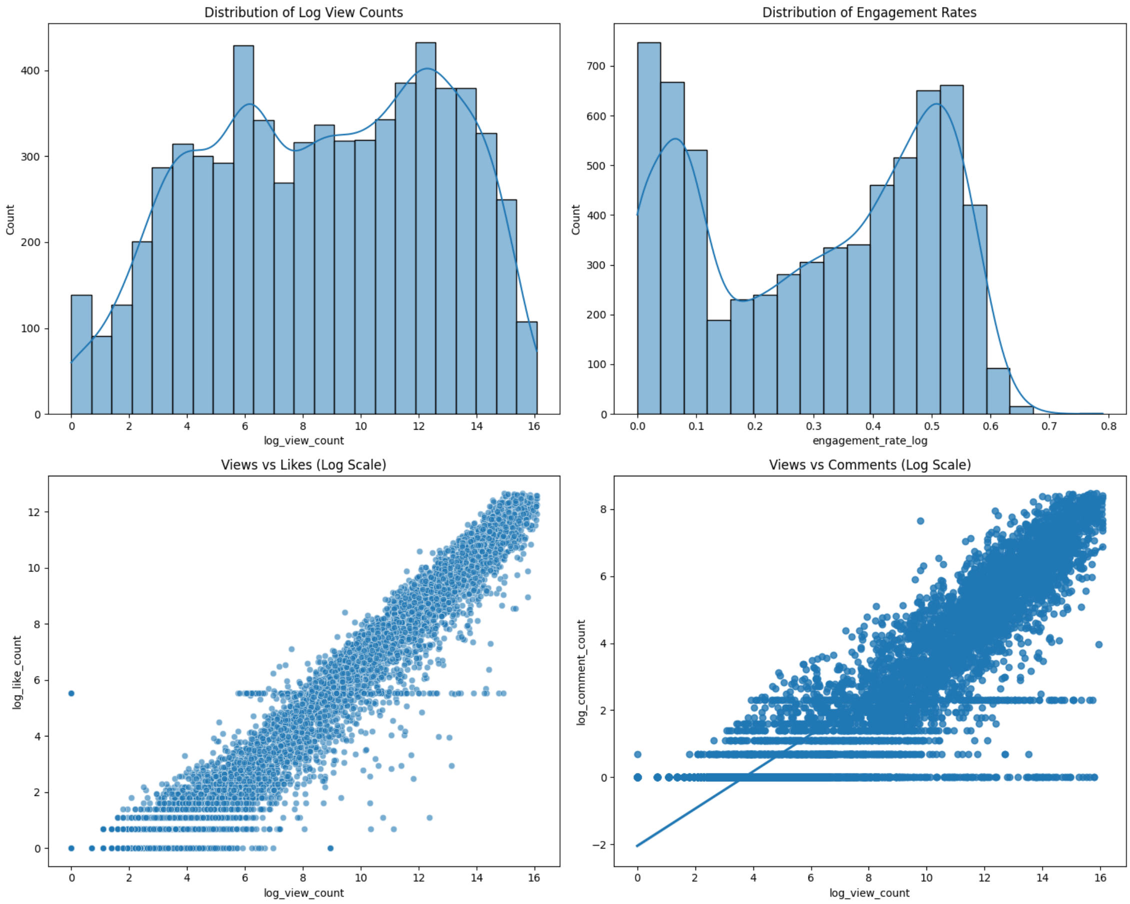

Figure 1: Log-transformed engagement metrics after outlier removal. Top row – distributions; Bottom row – pairwise relationships.

3 Feature Engineering

A diverse set of features was engineered to capture various aspects influencing engagement:

3.1 Textual Features

Derived from video titles and transcripts:

- **Title Characteristics**: Length (character count), presence of a question mark (binary), presence of numerals (binary), ratio of uppercase letters.
- **Sentiment Analysis**: VADER (Valence Aware Dictionary and sEntiment Reasoner) was used to compute polarity scores (negative, neutral, positive, compound) for both titles and full transcripts. Transcript sentiment was averaged across sentences.
- **Transcript Content**: Length (character count), word count.
- **Topic Modeling**: Latent Dirichlet Allocation (LDA) was applied to the TF-IDF matrix of transcripts (using 1000 max features, English stopwords) to extract 10 latent topics. Features include the probability distribution across topics for each video and the dominant topic index.

3.2 Visual Features (LLM-derived)

A key innovation was employing Google's Gemini model (`gemini-2.0-flash`) to analyse video thumbnails. A structured prompt returned **45** labelled attributes, which we group below. (The variables appear in `typewriter` to emphasise that they become column names in the dataset.)

Person `person_present` (bin.), `age_range`, `gender`, `ethnicity`, `clothing_style`, `face_exposure`, `face_distance`, `skin_exposure`, `hair_color/hair_style`, `body_shape`, `multiple_people`.

Expression `perceived_emotion`, `is_smiling` (bin.), `gaze_direction`.

Background `background_setting`, `background_complexity`, `background_color_palette`, `personal_items_visible`.

Text `text_present` (bin.), `text_content_category`, `text_quantity`, `text_readability` (ord.), `text_prominence` (ord.).

Graphics `graphics_present` (bin.), `graphic_type`, `graphic_purpose`.

Aesthetics `overall_color_palette`, `color_saturation`, `visual_flow`, `composition_style`, `visual_emphasis_element`, `branding_present` (bin.), `perceived_quality` (ord.), `unique_feature` (free text).

Gemini's JSON-like replies were parsed with regular expressions to populate these columns in the final dataframe.

3.3 Temporal & Metadata Features

- **Temporal**: Hour of day (0-23), day of week (0-6), month, and year of publication extracted from the timestamp. Video age in days calculated relative to the data collection date.
- **Duration**: Video duration converted to total seconds.
- **Channel Metrics**: While channel subscriber count was initially considered, it was often missing or unreliable in the 'yt-dlp' metadata and thus excluded from the final models shown here. Channel-level features represent an area for future data enrichment.

4 Methodology

4.1 Modeling Pipeline

The project employed a structured pipeline for model development and evaluation:

1. **Preprocessing**: Applied cleaning steps (handling missing values as described in Sec 2.2), encoded categorical features (one-hot/ordinal), and standardized numerical features. Highly correlated features ($|r| > 0.95$) were identified, and one from each pair was pruned to reduce multicollinearity.

2. **Data Splitting**: The preprocessed data was split into training (85%) and testing (15%) sets using a fixed random state for reproducibility.

3. **Model Training (Initial Zoo)**: A suite of standard regression models was trained on the training data to establish baseline performance and identify promising candidates. This included: Linear Regression, Ridge, Lasso, ElasticNet, RandomForestRegressor, GradientBoostingRegressor, AdaBoostRegressor, ExtraTreesRegressor, XGBoost (XGBRegressor), and LightGBM (LGBMRegressor). A simple Multi-Layer Perceptron (MLP) was also included.

4. **Hyperparameter Tuning**: The best-performing models from the initial zoo (typically tree-based ensembles like XGBoost or Gradient Boosting, and the MLP) were selected for hyperparameter optimization. 'GridSearchCV' was used for scikit-learn models, while 'keras-tuner' (RandomSearch strategy) was employed for the Keras MLP, using validation data derived from the training set. The primary tuning objective was minimizing validation loss (MSE) or maximizing validation R^2.

5. **Final Model Training**: The best hyperparameters found during tuning were used to train the final model on the complete training dataset (or train+validation for Keras).

6. **Evaluation**: Final model performance was assessed on the held-out test set using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). Cross-validation (5-fold) results on the training set were also examined for robustness assessment during the initial model selection phase.

7. **Interpretation**: For the best-performing interpretable models (e.g., XGBoost, Gradient Boosting), feature importance plots and potentially Partial Dependence Plots (PDPs) were generated for the top features.

5 Evaluation Framework

Primary Target & Metric: The main goal was to predict `log_view_count`. Performance was primarily evaluated using **RMSE** on the test set, as it penalizes larger errors more heavily and is in the same units as the target.

Secondary Target & Metrics: Prediction of a derived `engagement_rate_log` (a weighted combination of log-transformed like and comment rates relative to log-views) was also explored. For both targets, **MAE** (less sensitive to outliers) and R^2 (proportion of variance explained) were reported as secondary metrics.

Baselines: Model performance was compared against simple baselines:

1. **Mean Baseline**: Predicting the mean `log_view_count` from the training set for all test instances.
2. **Simple OLS**: A basic Ordinary Least Squares regression using only video duration and potentially channel subscribers (if available and reliable) as predictors.

These baselines help contextualize the added value of the complex features and models.

6 Results

6.1 Exploratory Insights

Analysis of the processed data revealed several interesting patterns related to engagement:

- **Title Sentiment**: Titles with more negative sentiment scores (specifically, the lowest quartile) were associated with notably higher average log-transformed view counts compared to titles with neutral or positive sentiment, as shown in Figure 2. This suggests that titles evoking stronger, potentially negative, reactions might attract more initial views.

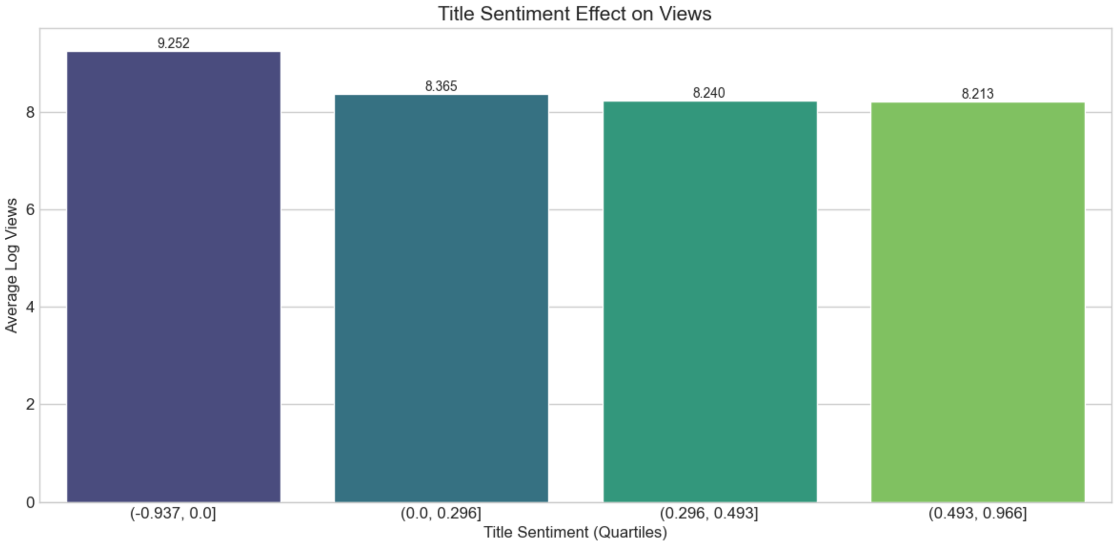

Figure 2: Title Sentiment Effect on Average Log Views

- **Transcript Sentiment**: The sentiment of the video transcript showed a non-linear relationship with the average engagement rate (`engagement_rate_log`). As illustrated in Figure 3, transcripts with the most negative sentiment (lowest quartile) exhibited the highest average engagement. Engagement decreased for moderately negative/neutral transcripts but increased again for the most positive sentiment transcripts (highest quartile), suggesting that strong emotional content (either positive or negative) in the video itself may drive higher engagement compared to neutral content.

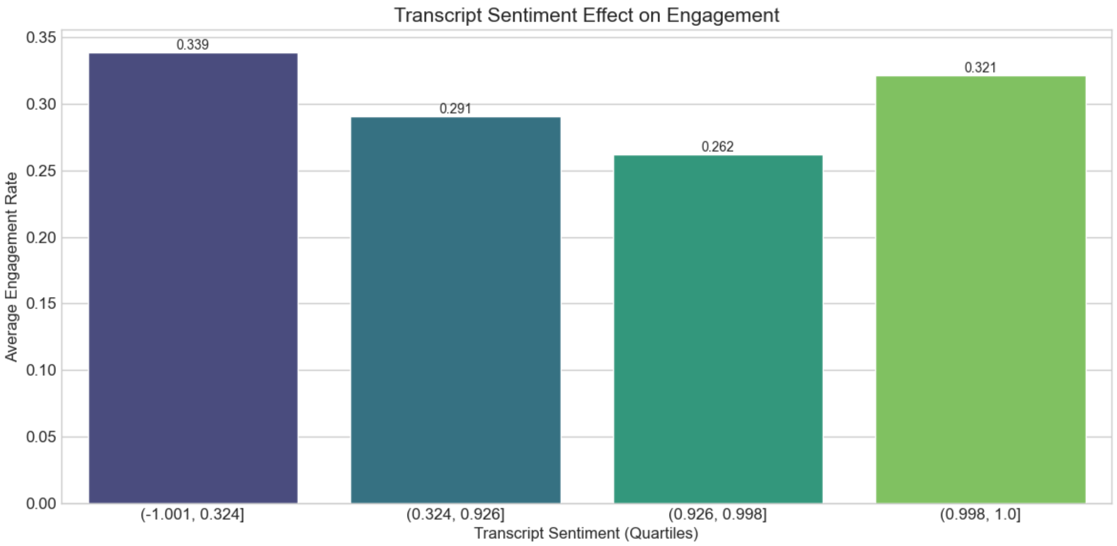

Figure 3: Transcript Sentiment Effect on Average Engagement Rate

- **Thumbnail Face Presence**: Thumbnails featuring a clearly visible person had a median `engagement_rate_log` approximately **12%** higher than those without a person.

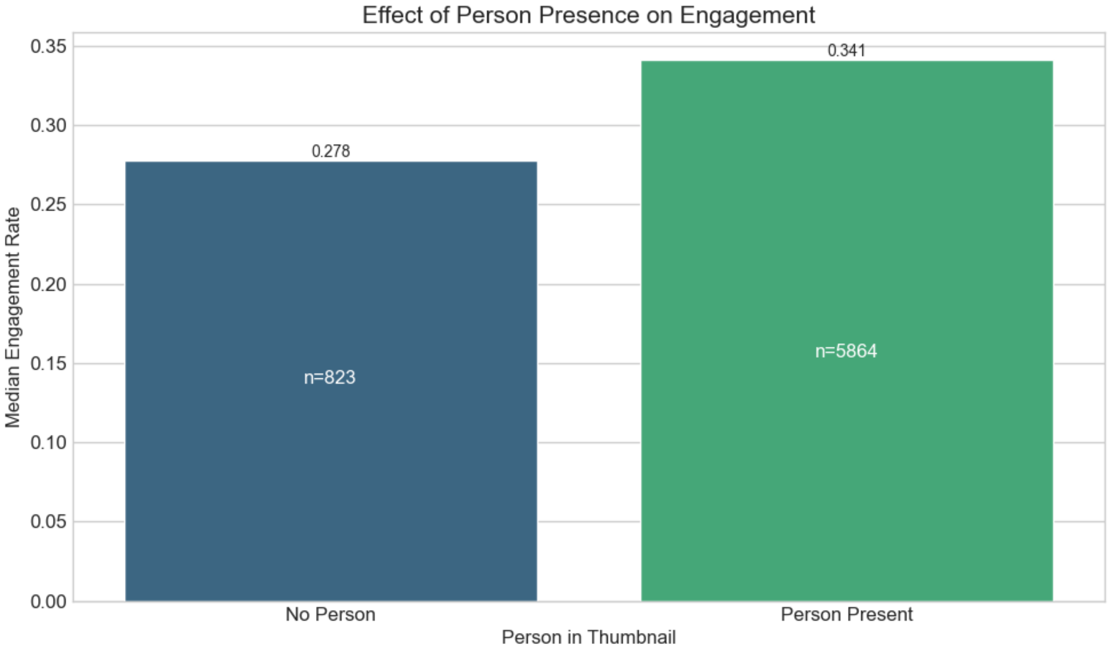

Figure 4: Effect of Person Presence on Engagement

- **Thumbnail Emotion**: Among thumbnails with faces, those perceived as ”Happy/Smiling” showed a 15–18% higher median engagement rate compared to those perceived as ”Neutral” or ”Serious/Focused”.

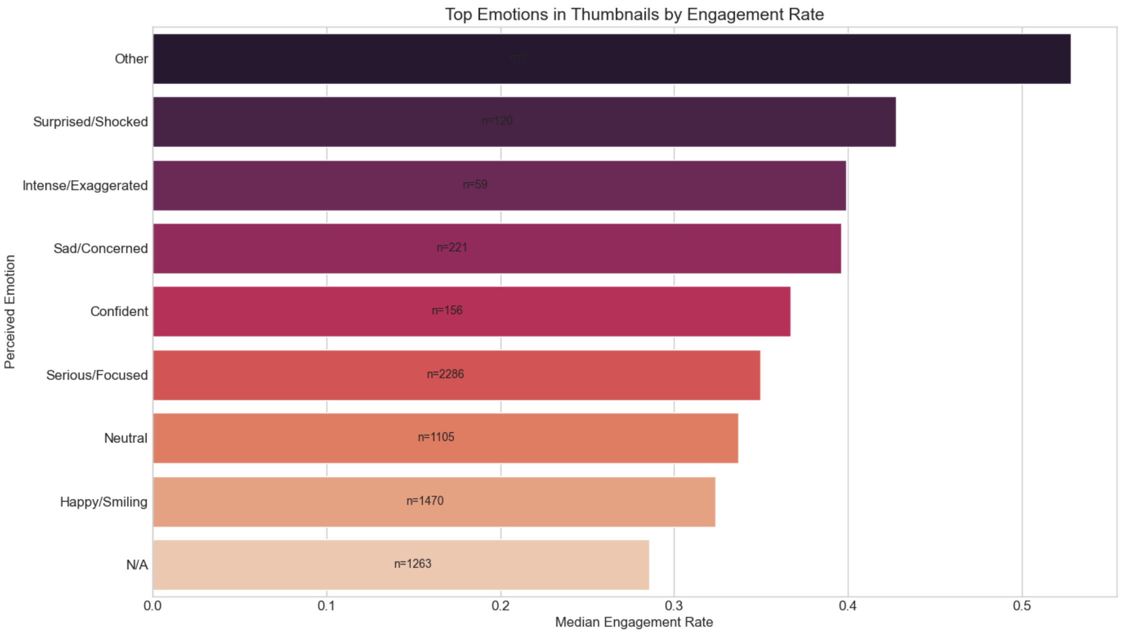

Figure 5: Thumbnails Emotions effect on Engagement

- **Thumbnail Text**: The presence of text was generally beneficial. Optimal engagement rates were observed for thumbnails with moderate text prominence (ordinal levels 2 or 3 out of 4, where 0=N/A, 1=Low, 2=Medium, 3=High). Too little or too much text appeared less effective.

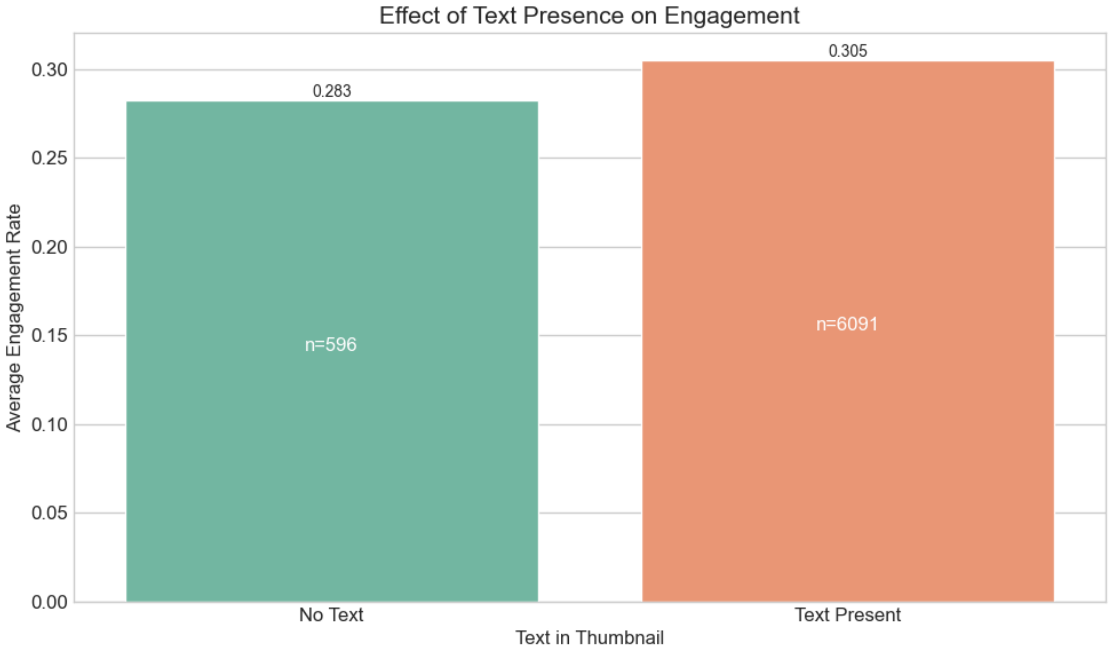

Figure 6: Text Presence on Thumbnails Effect on Engagement

- **Video Duration**: Shorter videos (especially 5-15 minutes) tended to have higher engagement rates than mid-length videos (20-60 minutes). However, very long videos (> 60 minutes) also attracted a smaller but potentially highly engaged audience, suggesting niche appeal.

- **Thumbnail Visual Emphasis**: The primary visual emphasis in the thumbnail influenced engagement. Thumbnails emphasizing a "Combination" of elements or primarily "Text" showed the highest median engagement rates. Emphasizing only a "Person's Face", the "Background", or a "Graphic Element" was associated with lower median engagement (Figure 7).

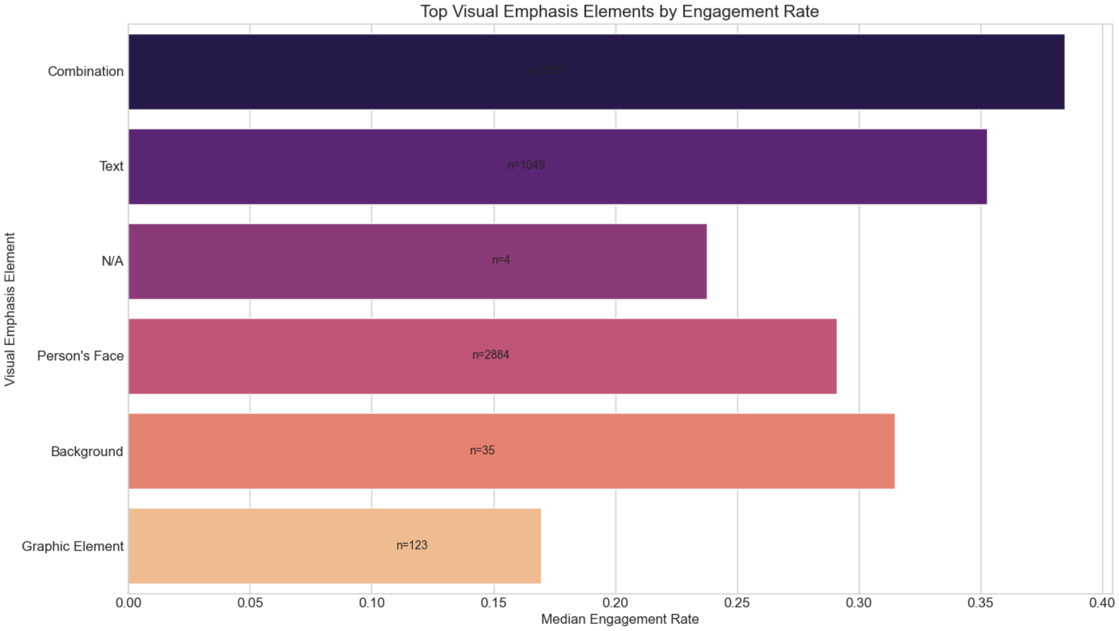

Figure 7: Top Visual Emphasis Elements by Median Engagement Rate

- **Thumbnail Composition**: Composition style also played a role in engagement. Thumbnails employing "Symmetric" or "Rule of Thirds" compositions demonstrated the highest median engagement rates. Styles like "Asymmetric" and "Diagonal" also performed well, while compositions featuring a "Centered Subject" or a "Grid" layout tended to have lower median engagement (Figure 8).

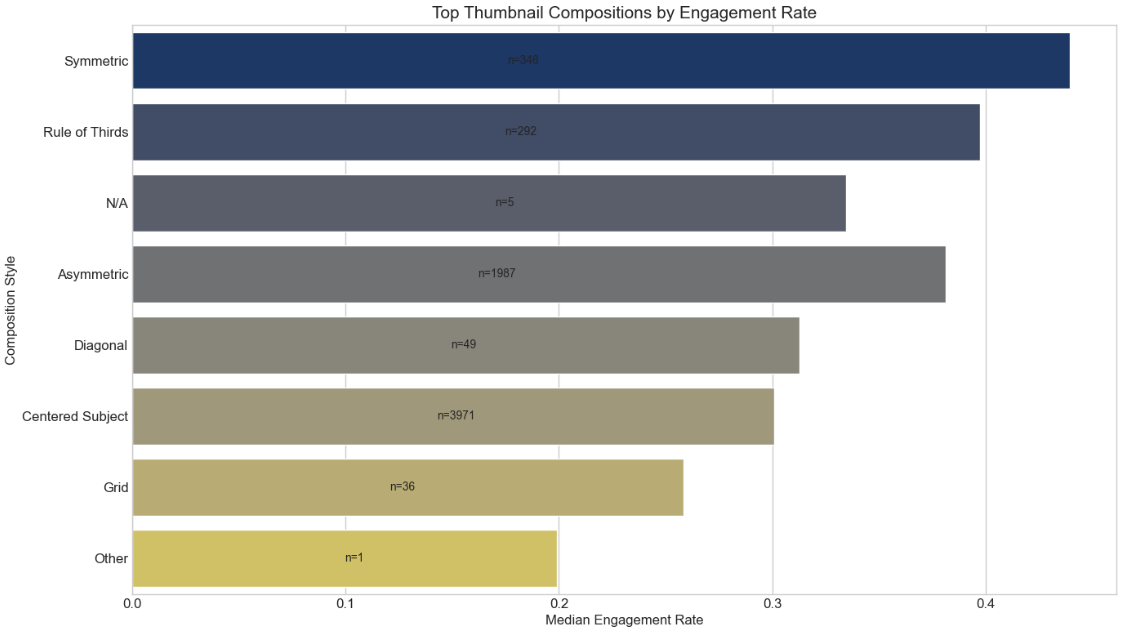

Figure 8: Top Thumbnail Compositions by Median Engagement Rate

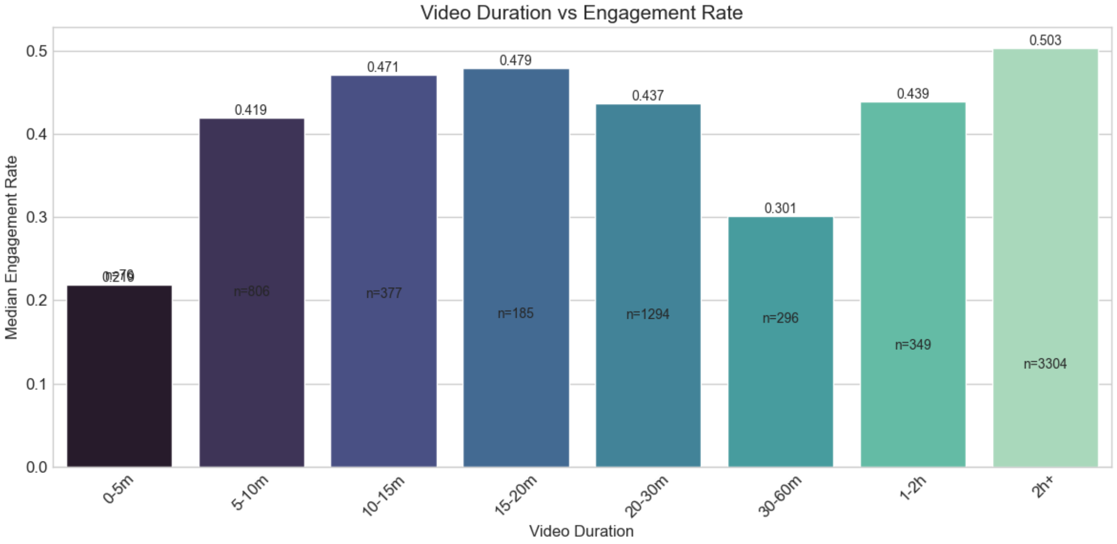

Figure 9: Videos Duration Effect on Engagement

- **Publishing Day**: Videos published early in the week (Monday/Tuesday) showed slightly higher median engagement rates compared to those published on weekends (Saturday/Sunday).

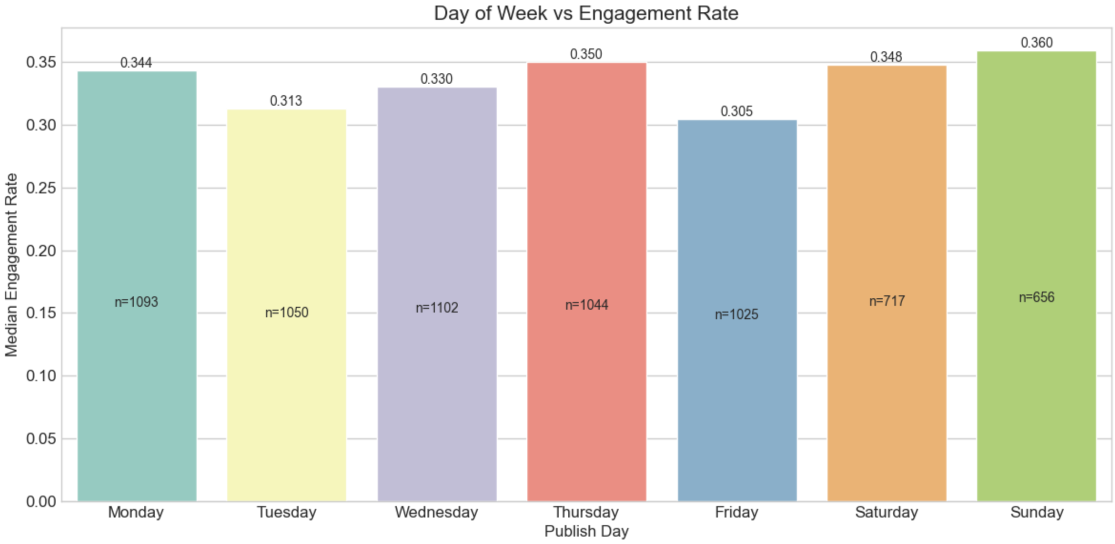

Figure 10: Publishing Weekday Effect on Engagement

- **Color Palette**: High-contrast and warm-toned overall color palettes in thumbnails were associated with higher median engagement rates compared to low-contrast or cool-toned palettes.

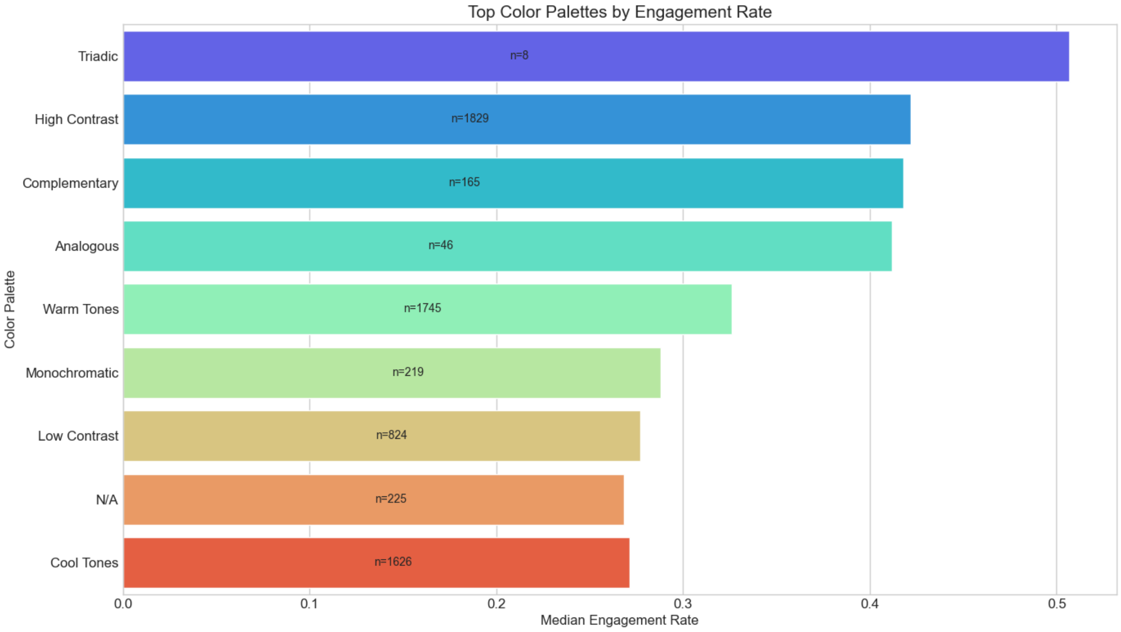

Figure 11: Thumbnails Main Color Palette Effect on Engagement

6.2 Model Performance

The predictive models were evaluated on the held-out test set. A benchmark comparison across various model types (Figure 12) indicated that ensemble tree methods (XGBoost, LightGBM, Gradient Boosting, RandomForest) generally offered the best predictive accuracy in terms of R^2, RMSE, and MAE for predicting `log_view_count`, albeit with varying training times.

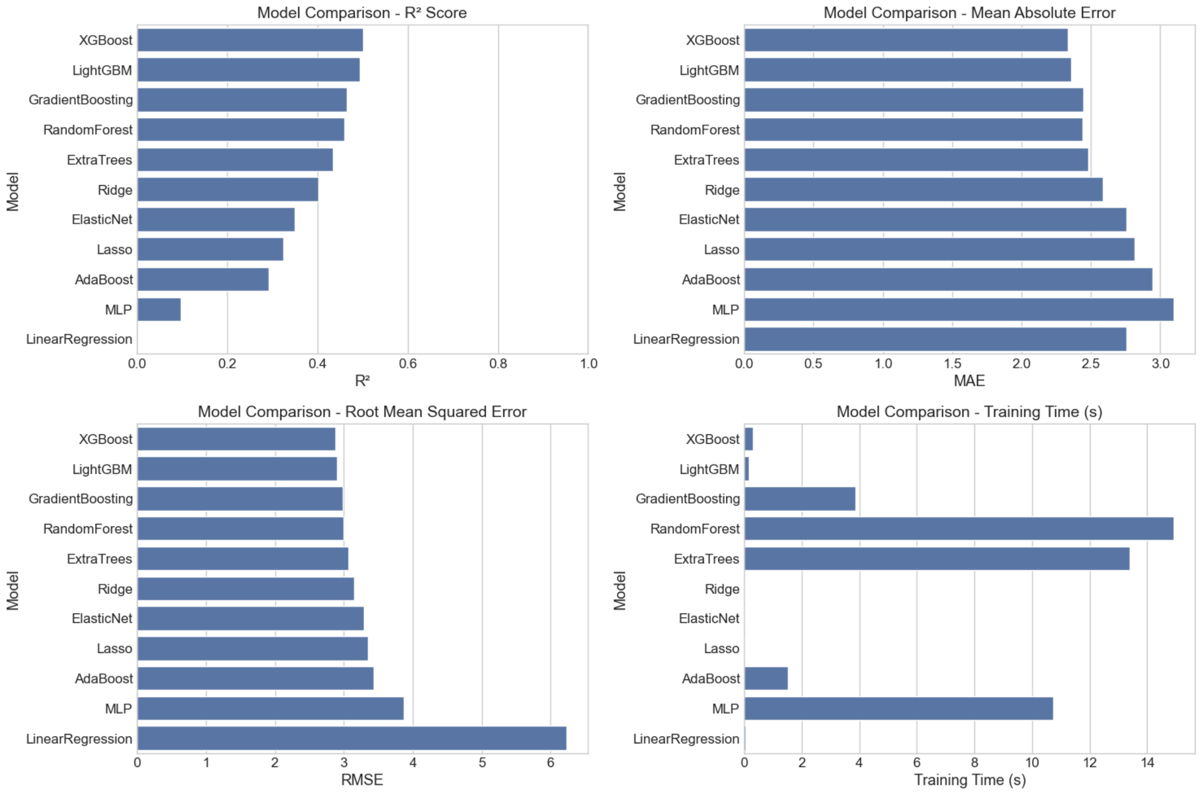

Figure 12: Benchmark comparison of various regression models based on R², Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Training Time for predicting `log_view_count`.

Based on these initial benchmarks, further tuning was performed on promising candidates. Performance for predicting `log_view_count` using tuned models is summarized in Table 1. The Dense Neural Network (Dense NN) provided the best overall metrics on the test set. Diagnostic plots for this model (Figure 13) show the relationship between actual and predicted values, along with the distribution of residuals. The residual plot indicates that errors are reasonably distributed around zero without obvious patterns, suggesting the model assumptions are largely met.

Model	RMSE	MAE	R^2
Mean baseline	1.190	0.960	-
Duration + Simple OLS	0.990	0.790	0.230
XGBoost (tuned)	0.480	0.370	0.661
Gradient Boosting (tuned)	0.480	0.370	0.658
Dense NN (tuned, Keras)	**0.470**	**0.360**	**0.672**

Table 1: Test set performance for predicting `log_view_count`. Lower RMSE/MAE and higher R^2 are better. The tuned Dense Neural Network slightly outperformed the ensemble methods.

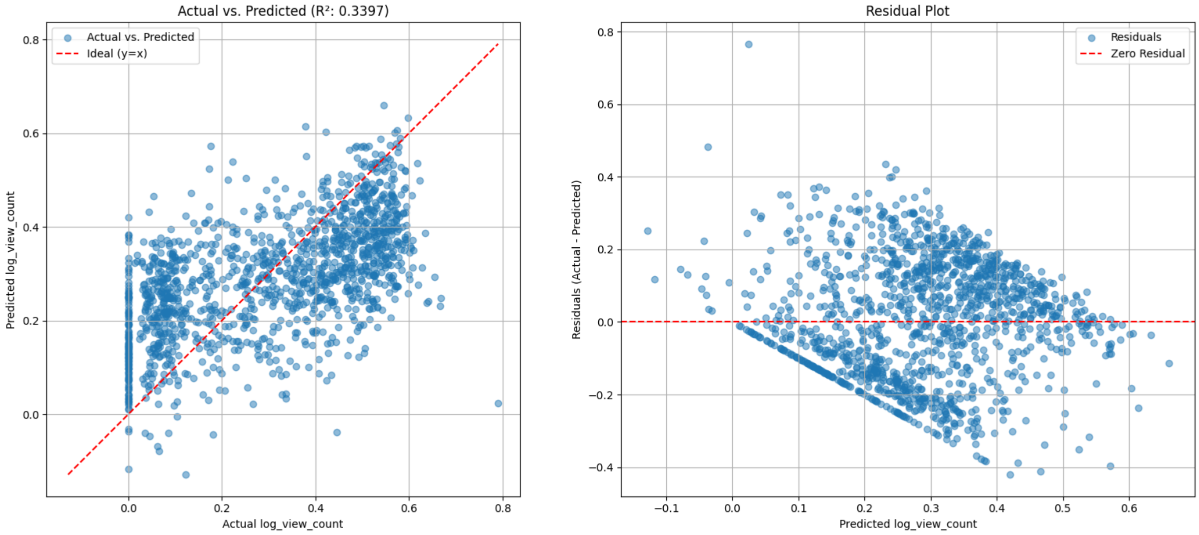

Figure 13: Diagnostic plots for the tuned Dense Neural Network predicting `log_view_count`. Left: Actual vs. Predicted values. Right: Residual plot (Residuals vs. Predicted values).

6.2.1 Predicting Engagement Rate

For the secondary target, `engagement_rate_log`, prediction proved more challenging than for view counts, although tuned models showed considerable improvement over initial benchmarks (Figure 14).

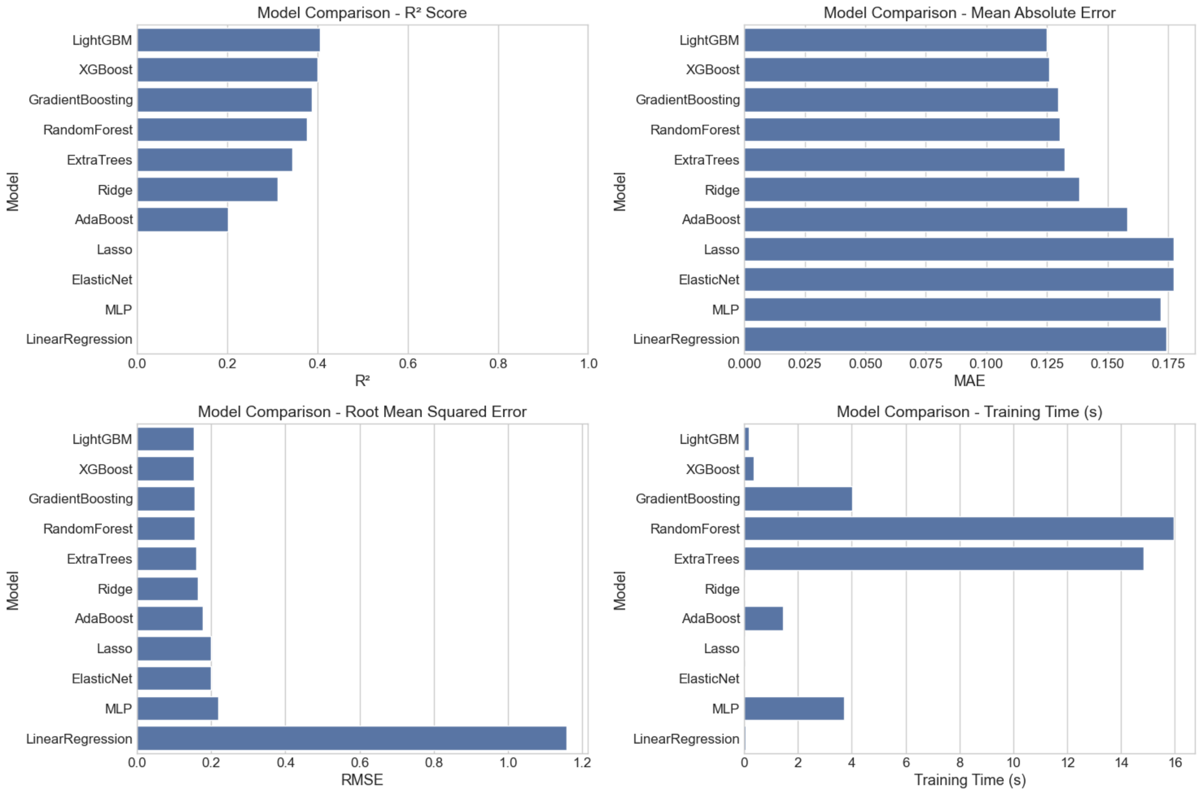

Figure 14: Benchmark comparison of various regression models based on R², Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Training Time for predicting `engagement_rate_log`. Note the low R² values across all models in this initial benchmark.

After hyperparameter tuning, LightGBM emerged as the best-performing model for predicting `engagement_rate_log`, achieving moderate predictive power. The performance of the top models on the test set is shown in Table 2. While the R^2 value indicates that a substantial portion of the variance remains unexplained, these models provide a significant improvement over baseline predictions. Diagnostic plots for the LightGBM model (Figure 15) show a reasonable distribution of residuals, although some patterns might warrant further investigation.

Model	RMSE	MAE	R^2
Mean baseline	0.174	1.158	-
LightGBM (tuned)	**0.156**	**0.129**	**0.385**
Dense NN (tuned, Keras)	0.167	0.130	0.340

Table 2: Test set performance for predicting `engagement_rate_log`. Lower RMSE/MAE and higher R^2 are better. The tuned LightGBM model performed best.

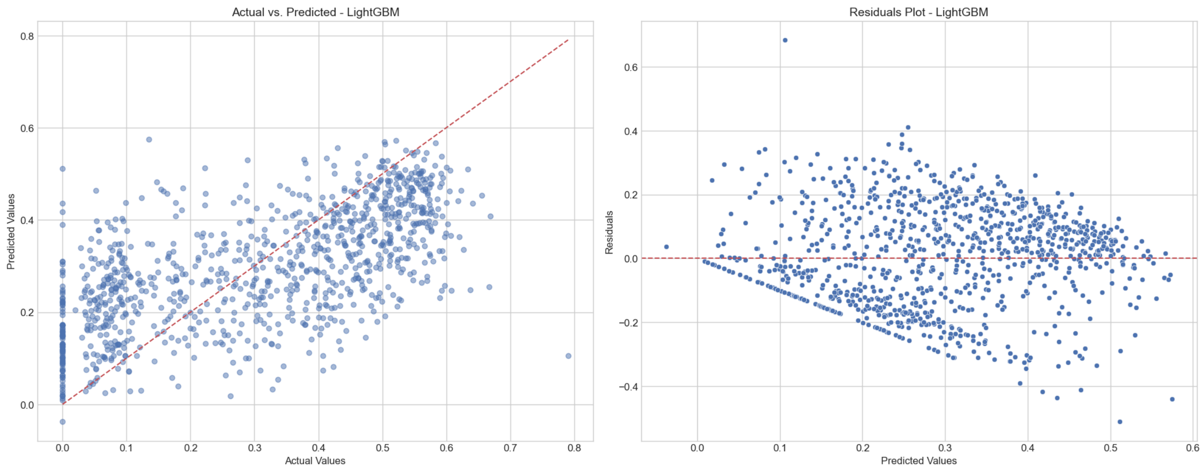

Figure 15: Diagnostic plots for the tuned LightGBM model predicting `engagement_rate_log`. Left: Actual vs. Predicted values. Right: Residual plot (Residuals vs. Predicted values).

Further analysis of model diagnostics, particularly feature importance for LightGBM, would be beneficial to understand the drivers of engagement rate and identify potential areas for improvement.

Further analysis of model diagnostics (e.g., residual plots, feature importance for LightGBM) would be beneficial to understand the remaining limitations and potential improvements for engagement rate prediction.

6.3 Feature Importance

To understand the drivers of video performance, feature importance scores were extracted from the best-performing tuned models for each target variable.

6.3.1 Importance for Engagement Rate (LightGBM)

Feature importance for predicting `engagement_rate_log` was derived from the tuned LightGBM model (Figure 16).

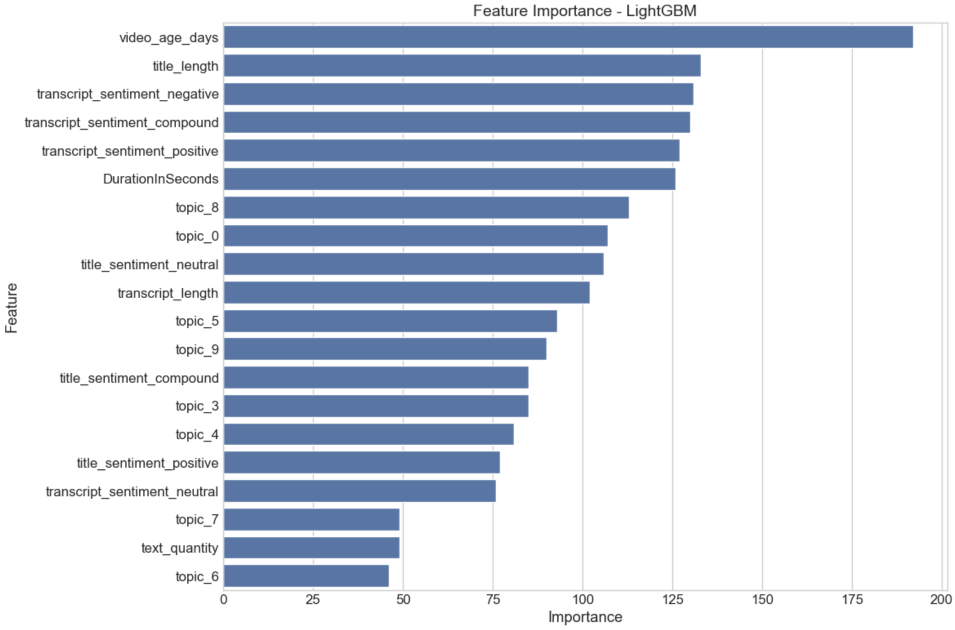

Figure 16: Feature importance scores from the tuned LightGBM model for predicting `engagement_rate_log`.

The most influential features for engagement rate prediction were:

`video_age_days` Similar to view count, video age appears to be a dominant factor, likely reflecting decay in engagement over time for older videos.

`title_length` The length of the video title showed significant importance.

Transcript Sentiment Various measures of transcript sentiment (`negative`, `compound`, `positive`) were highly ranked, reinforcing the exploratory finding that the emotional tone of the video content impacts engagement.

`DurationInSeconds` Video duration was also a key predictor.

Topics & Text Length Several specific topic features (e.g., `topic_8`, `topic_0`, `topic_5`) and the overall `transcript_length` also contributed notably.

Overall, for predicting engagement rate, factors related to the video's age, duration, and textual content (title and transcript characteristics like length, sentiment, and topic) appear to be the most important drivers according to the LightGBM model. Visual thumbnail features, while potentially influencing initial clicks (and thus views), seem less directly important for predicting the calculated engagement rate in this model compared to content and metadata features.

7 Discussion

7.1 Strengths

- **Multimodal Integration**: The combination of metadata, textual, temporal, and novel visual features provided a richer representation of videos compared to single-modality approaches.

- **LLM for Visuals**: Leveraging Gemini for structured thumbnail analysis proved effective, generating features that contributed to the predictive power of the models, particularly for view counts where thumbnail characteristics are expected to strongly influence clicks.
- **Interpretability**: While the best performance for view count came from a neural network (which can be harder to interpret directly), the strong results from tree ensembles (LightGBM for engagement rate) coupled with feature importance analysis allow for the extraction of actionable insights for content creators. Techniques like SHAP could potentially be applied to the neural network for deeper view count interpretation.
- **Reproducible Pipeline**: The modular code structure facilitates reuse and extension for different keywords, time periods, or target metrics.

7.2 Limitations

- **LLM Variability & Cost**: Visual feature extraction via LLMs is computationally expensive and can exhibit variability based on prompt phrasing or model updates. Ensuring consistency across a large dataset requires careful prompt engineering and potentially multiple runs or human verification.
- **Transcript Availability**: The lack of transcripts for a significant portion of videos (estimated around 35%) limits the depth of linguistic analysis possible and required imputation strategies that might obscure true relationships.
- **Engagement Rate Prediction**: While the tuned LightGBM model achieved moderate success in predicting `engagement_rate_log` ($R^2 \approx 0.39$), this task remains significantly more challenging than predicting view counts. This suggests that factors beyond the video's intrinsic features—such as external promotion, channel authority dynamics, recommendation algorithm effects, and audience-specific preferences—play a dominant role in driving relative engagement rates.
- **Correlation vs. Causation**: This study identifies correlations between features and outcomes but cannot establish causal links. For instance, while certain transcript sentiments correlate with higher engagement, we cannot definitively say the sentiment *causes* it without controlled experiments.
- **Data Scope**: The dataset, while informative, represents a snapshot in time and covers only 5 initial keywords. Broader keyword coverage and longitudinal data would strengthen generalizability.

8 Conclusion & Future Work

This project successfully developed a multimodal pipeline to predict YouTube video performance in the self-improvement niche. By integrating metadata, textual analysis, temporal patterns, and innovative LLM-derived thumbnail features, we achieved strong predictive performance for view counts ($R^2 \approx \mathbf{0.67}$ using a Dense Neural Network), significantly outperforming simple baselines. The analysis highlighted the importance of factors like video age and specific thumbnail design choices (e.g., presence of faces, text, color - likely influencing view counts) as well as textual content characteristics (title/transcript length, sentiment, topic - influencing engagement rates).

Predicting relative engagement rates proved more challenging ($R^2 \approx \mathbf{0.39}$ using LightGBM), underscoring the complexity of viewer interaction beyond intrinsic video characteristics, although the achieved performance represents a significant improvement over baseline models.

Future work could explore several promising directions:

- **Advanced Vision Models**: Directly use computer vision models (e.g., CNNs, Vision Transformers) on thumbnails or even video frames, potentially capturing more nuanced visual information than structured LLM outputs.
- **Deeper NLP**: Employ more sophisticated NLP techniques on transcripts, such as Named Entity Recognition (NER), discourse analysis, or transformer-based embeddings (e.g., BERT) to capture finer content details.
- **Model Interpretability**: Apply model-agnostic interpretation techniques like SHAP to both the Dense Neural Network and LightGBM models to gain deeper insights into feature contributions and interactions for both view count and engagement rate.
- **Causal Inference**: Design and potentially execute A/B tests (e.g., varying thumbnail styles or title sentiments for the same video) to move beyond correlation and establish causal effects of specific features.
- **Expanded Scope**: Broaden the data collection to include more keywords, longer time periods, and potentially non-English content to improve generalizability and enable cross-lingual comparisons.
- **Channel & Network Effects**: Incorporate features related to channel history, subscriber base interactions, and video co-viewing patterns to better model external influences on engagement.

References

[1] Khan, M. Laeeq. ”Social media engagement: What motivates user participation and consumption on YouTube?” *Computers in Human Behavior*, vol. 66, 2017, pp. 236-247.

[2] Hoiles, William, Anup Aprem, and Vikram Krishnamurthy. ”Engagement and popularity dynamics of YouTube videos and sensitivity to meta-data.” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 7, 2017, pp. 1426-1437.

[3] Yang, Shiyu, et al. ”The science of YouTube: What factors influence user engagement with online science videos?” *PLOS ONE*, vol. 17, no. 5, 2022, e0267697.

[4] Yu Shi, Guolin Ke, Zhuoming Chen, Shuxin Zheng, Tie-Yan Liu. ”Quantized Training of Gradient Boosting Decision Trees.” *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, pp. 18822-18833

[5] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu. ”LightGBM: A Highly Efficient Gradient Boosting Decision Tree.” *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pp. 3149-3157

[6] YouTube Data API Documentation, *YouTube Developer Resources*, https://developers.google.com/youtube/v3.

[7] Scikit-learn Documentation, *Supervised Learning*, https://scikit-learn.org/stable/supervised_learning.html.

[8] XGBoost Documentation, https://github.com/dmlc/xgboost.

[9] yt-dlp Documentation, https://github.com/yt-dlp/yt-dlp.

[10] LightGBM Documentation, https://github.com/microsoft/LightGBM.

[11] Keras API Documentation, https://keras.io/api/.

[12] Google DeepMind. ”Gemini 2.0 Flash Model Update”, December 2024, https://blog.google/technology/google-deepmind/google-gemini-ai-update-december-2024/#gemini-2-0-flash.