YouTube Title Recommender

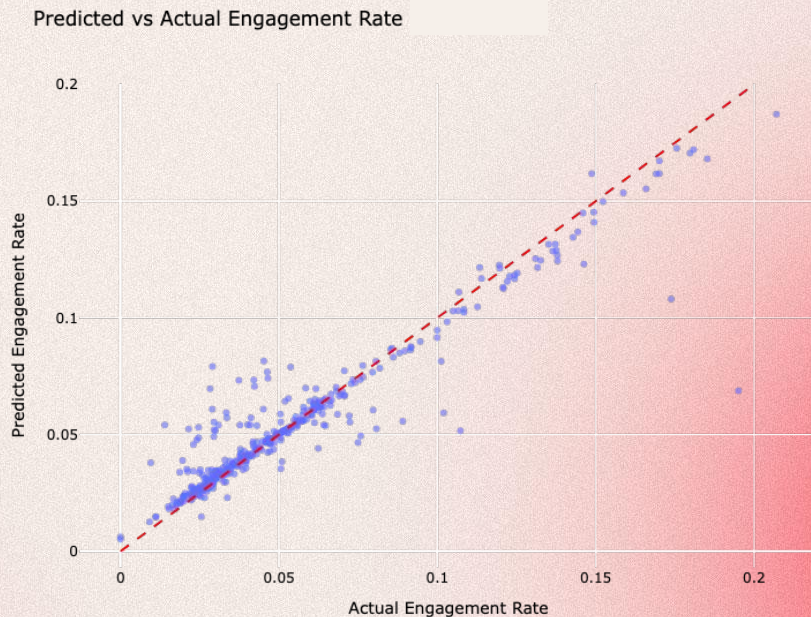UCLA STATS418
STEPHANIE LU

JUNE 2025

Data Collection

- Queried 20 lifestyle/wellness YouTube channels using YouTube Data API v3
- Pulled each channel's 50 most recent videos since January 1st, 2025
- Filtered data to exclude content likely to be sponsored based on keywords
- Final dataset: 402 videos with view, like, and comment counts
- Created new feature engagement_rate
  - (likes + comments) / views
  - Industry standard for evaluating creator fit with brand campaigns

Modeling

Predicting video engagement based on title text
- Cleaned title strings using regex and lowercasing
- Vectorized titles with TF-IDF
- Trained a ridge regression model to predict engagement rate
  - Used gridsearch to tune learning rate
  - RMSE = 0.03
  - Range of engagement rate: [0, 0.21]

Predicted vs Actual Engagement Rate

The App

-
- Input is vectorized, put into the model, which then uses cosine similarity to pull similar existing titles in the dataset
- Returns top 4 similar titles with high true engagement for inspiration

YouTube Title Recommender 🔗

Enter your video concept to discover trending, high-engagement tiles from popular lifestyle creators.

Enter a video title idea:

morning routine

Top Similar High-Engagement Titles

	Title	View Count	Engagement Rate
0	my 3am morning routine	263243	3.3%
1	Realistic 5AM morning routine ☕ HOW to become a morning girl as a ex-nightowl	62039	3.3%
2	UNI VLOG 📓 6AM morning routine, studying in the library, cleaning my apartment, skincare routine	120022	2.5%
3	What I Never Skip in My Slow Morning Routine	29622	2.3%

Next Steps

- Distinguish between long-form and short-form videos
- Use word embeddings e.g. Word2Vec to capture more semantic similarity
- Add interactibility to app e.g. optimize for views instead of engagement
- Expand to other niches + allow user to select one in-app