

POWERPOST: USING MACHINE LEARNING ADVERSERIALLY WITH GENERATIVE AI TO OPTIMIZE LINKEDIN POSTS



CARMEL SOCEANU, IDAN HOROWITZ, LIAN FICHMAN, MATAN BIRNBOIM

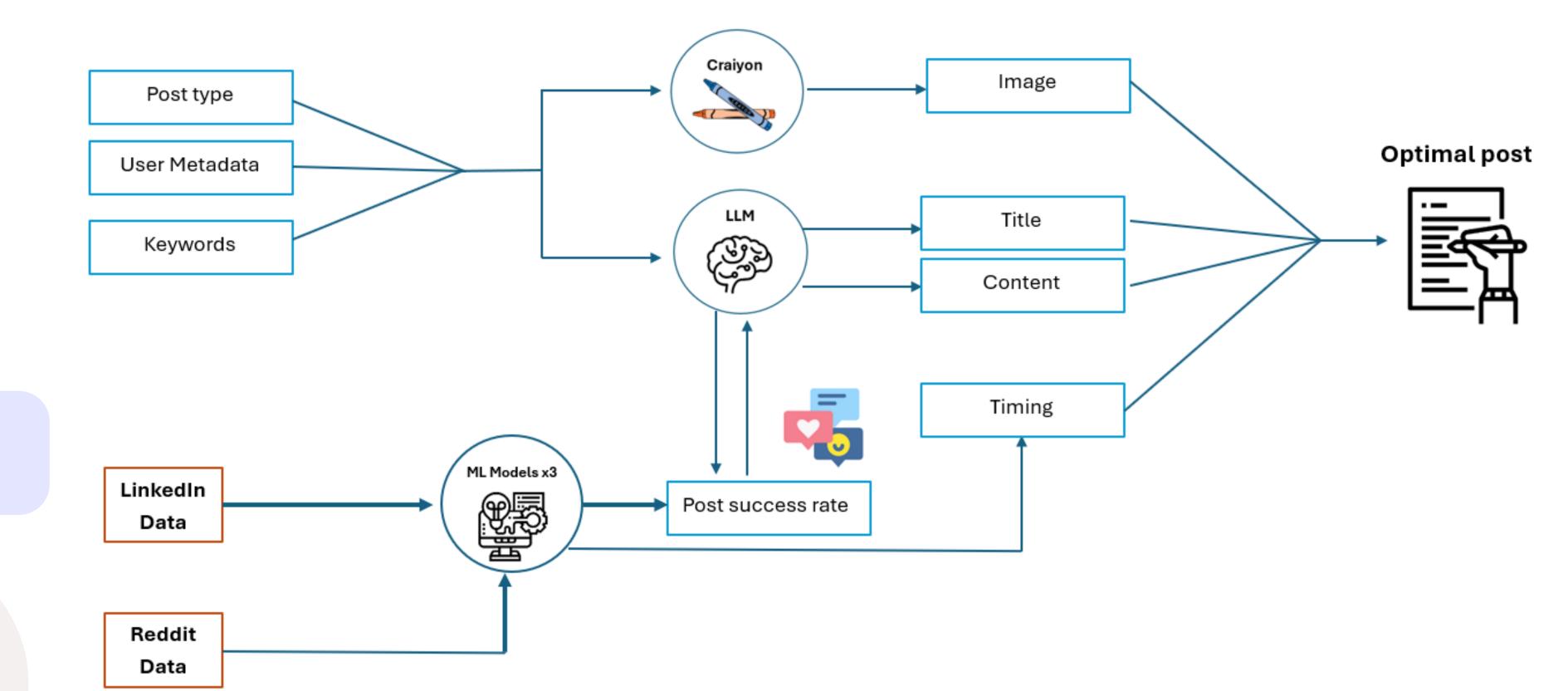
BACKGROUND

GOALS: CREATE A TOOL FOR
LINKEDIN USERS TO
ENHANCE ENGAGEMENT
THROUGH TIMING
OPTIMIZATION, TITLE
GENERATION, CONTENT
EVALUATION, AND IMAGE
PRODUCTION.

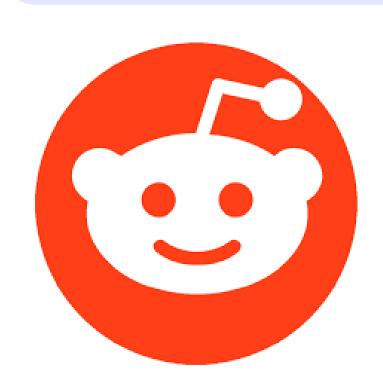
AIMS: DISCOVER THE FACTORS THAT CREATE ENGAGING, POPULAR POSTS AND HOW CAN WE ASSIST THE EVERYDAY USER IN IMPLEMENTING THEM INTO THEIR CONTENT?

METHODS: USE ML TO
LEARN OPTIMAL POST
PARAMETERS AND IN AN
'ADVERSERIAL' STYLE
UTILISE AN LLM TO
GENERATE THE PERFECT
POST

ARCHITECTURE



DATA



SCRAPED
DATA FROM
REDDIT:
22000 ROWS

REDDIT
READY DATA:
133000
ROWS

PROVIDED LINKEDIN DATASETS

NORMALISING FOLLOWER SCORE

GOAL: NORMALISE FOLLOWER COUNT TO REMOVE NOISE AND ATTRIBUTE VALUE TO POST SUCCESS ALONE

$$NewFollowerValue(x) = \begin{cases} \frac{x-12}{Baseline}, & x > 12 \\ 0, & otherwise \end{cases} Baseline = ModelR^2 \cdot \left(\sum_{i \in features} Importance_i \cdot \mu_i\right)$$

IN OUR CASE:

MODEL USED: RANDOM FOREST REGRESSION

MODEL R^2 = 0.09

SIGNIFICANT FEATURES: COMPANY, POSITION, EDUCATION

WORKFLOW

Data Collection:

- Scraped data from Reddit, capturing post score, timing, title, and subreddit and joined to pre-existing dataset.
- Extracted both company & profile LinkedIn data, organizing each row to represent a post.

Optimal Posting Time:

- Analyzed Reddit data to determine the optimal time for posting.
- Utilized timing patterns across categories to suggest the best time.

Data Preprocessing:

- Filtered LinkedIn dataset, removing rows with missing values in post content and total likes.
- Performed data scaling, imputation and encoding.
- Normalised follower value to remove confounders and obtain baselines.

Jata Preprocessing:

- Feature Engineering & Model Training:
 Generated several descriptive features based on profile data.
 - Employed Word2Vec and TF-IDF to extract features from post content.
 - Predicted total likes based on content using Random Forest & Linear Regression models.
 - Obtained feature importances and optimal post parameters.

Generative Model Integration:

- Use of LLM to create image generation prompt.
- Use of LLM to create post title & text.
- Predict success of generated post using ML model above and keep returning in adversarial style until perfect post obtained.

RESULTS

category_index

type_index

about features

company_size

0.005

0.015

0.010

0.020

Feature Importance

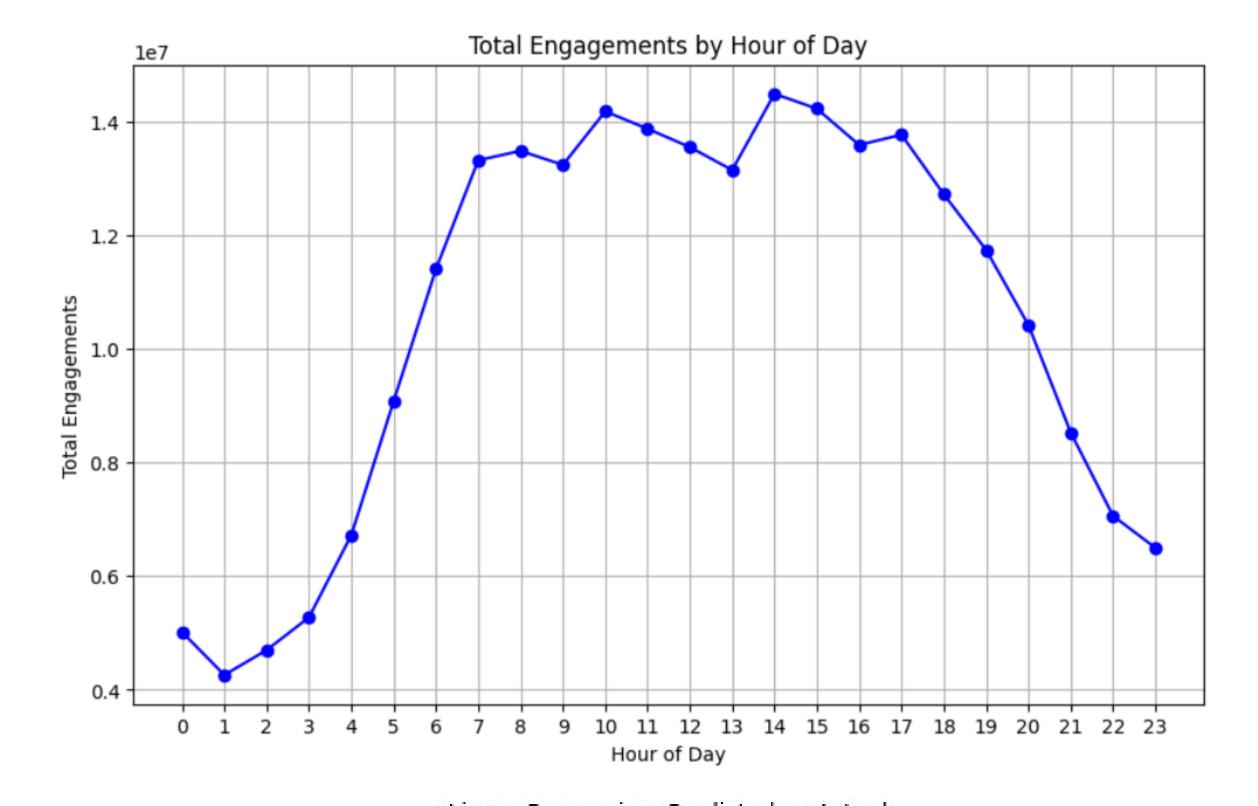
0.025

0.030

0.035

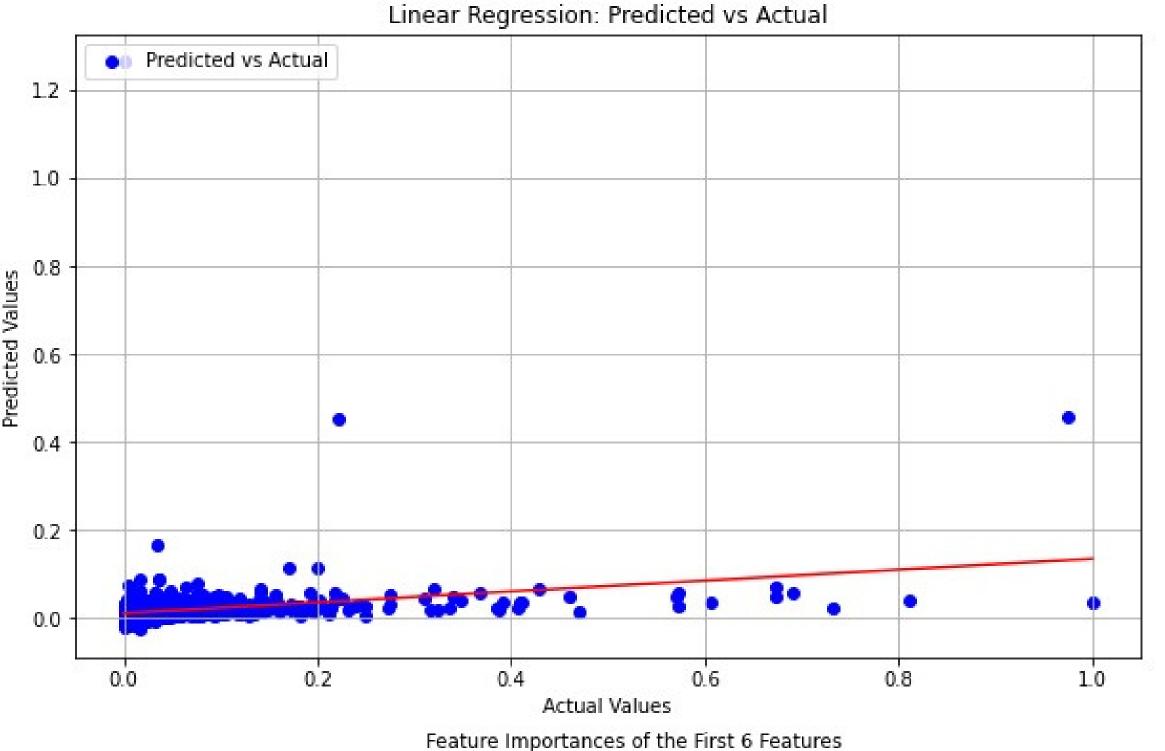
0.040

logo



OPTIMAL
POSTING
TIME: REDDIT

TIME CLASS.
ACCURACY:
0.68



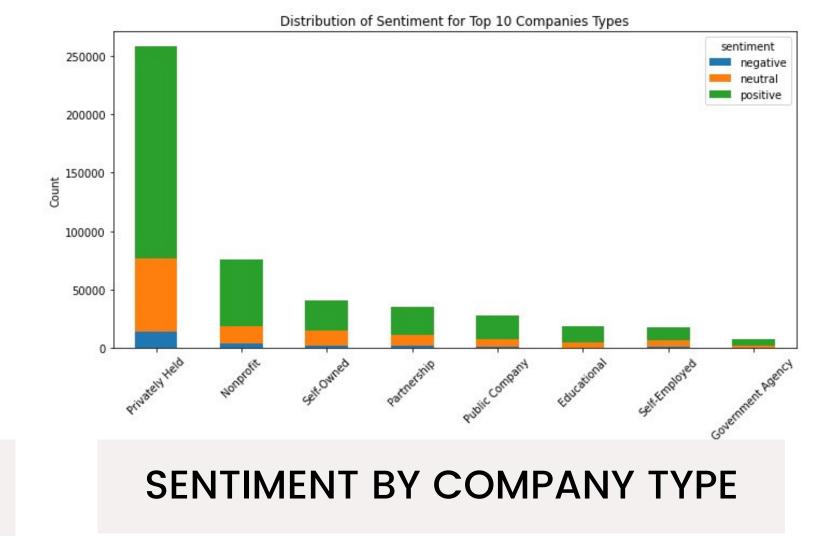
REGRESSION
FIT: POST
SUCCESS
PREDICTION

LINKEDIN
RESULTS
RMSE: 0.025
R^2: 0.324

RANDOM FOREST FEATURE IMPORTANCE

ANALYSIS





MOST ENGAGING WORDS: REDDIT