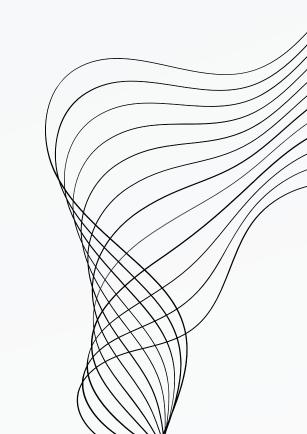


DEVESH SHAH, VEDIKA AGARWAL, YASH SRIVASTAVA

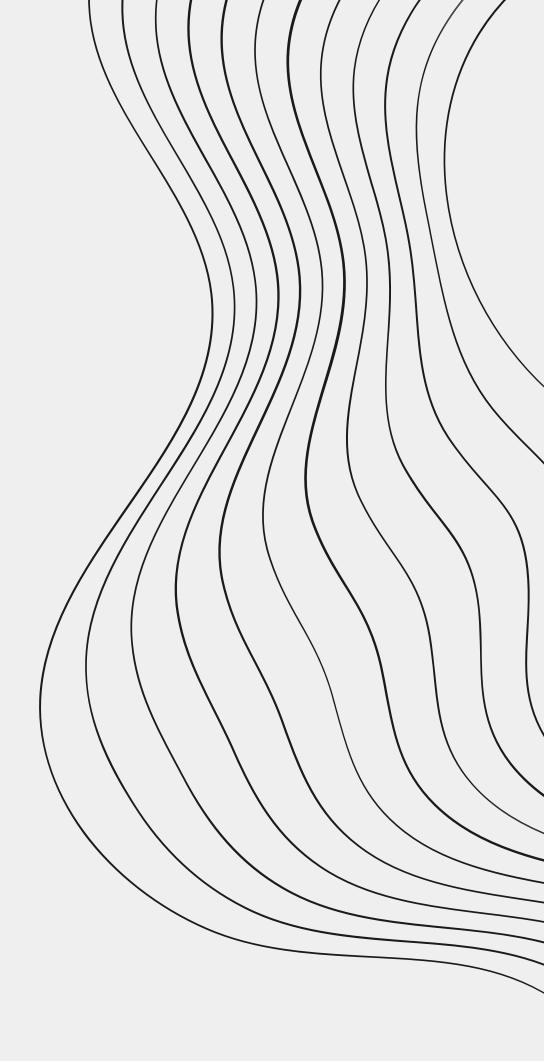


THE PROBLEM

Social Media is a powerful way to reach prospective customers and boost marketing. However, navigating the complex landscape of social media is a difficult task and leaves much to chance.

THE SOLUTION

Creating a Social Media Simulator that generates realistic social interactions that may emerge as a response to posts by businesses and the ability to fine-tune them to perform better.



3.2 BILLION

Active Social Media Users Worldwide

MOTIVATION

- In a survey of 3,700 marketers, 96% marketers said they use social media for marketing and 91% said they struggle to answer the question concerning the best ways to engage their target audiences on social media platforms.
- In a survey of more than 1,500 marketers, 72% stated that their top social media priority is to create more engaging content, and their second highest priority (65%) is to improve their understanding of what content is effective
- Social Media is a pivotal marketing platform, yet its dynamic and unpredictable nature poses challenges for businesses.

Pulizzi, J., & Handley, A. (2015). B2B Content Marketing: 2016 Benchmarks, Budgets and Trends — North America. Retrieved June 8, 2016, from http://contentmarketinginstitute.com/research/content- marketing-research/

IMPACT OF SOLUTION

Large Audience

Tighter knit communities may react very differently to posts than a large-scale audience, so the ability to test it out on the latter is an extremely powerful tool.

Design Choices

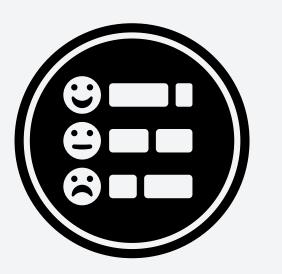
Designers of posts can reiterate more concretely on
their design for better
engagement by testing
them using the simulator
and receiving feedback

Sample Space of Responses

The simulator can not predict the complexities of social behaviour with 100% accuracy, but can generate a wide range of possible responses that may not have been anticipated.









PAPER 1

Social Simulacra: Creating Populated Prototypes for Social Computing Systems

Joon Sung Park Stanford University Stanford, USA joonspk@stanford.edu

Meredith Ringel Morris Google Research Seattle, WA, USA merrie@google.com Lindsay Popowski Stanford University Stanford, USA popowski@stanford.edu

Percy Liang Stanford University Stanford, USA pliang@cs.stanford.edu Carrie J. Cai Google Research Mountain View, CA, USA cjcai@google.com

Michael S. Bernstein Stanford University Stanford, USA msb@cs.stanford.edu

ABSTRACT

Social computing prototypes probe the social behaviors that may arise in an envisioned system design. This prototyping practice is currently limited to recruiting small groups of people. Unfortunately, many challenges do not arise until a system is populated at a larger scale. Can a designer understand how a social system might behave when populated, and make adjustments to the design before the system falls prey to such challenges? We introduce social simulacra, a prototyping technique that generates a breadth of realistic social interactions that may emerge when a social computing system is populated. Social simulacra take as input the designer's description of a community's design-goal, rules, and member personas- and produce as output an instance of that design with simulated behavior, including posts, replies, and anti-social behaviors. We demonstrate that social simulacra shift the behaviors that they generate appropriately in response to design changes, and that they enable exploration of "what if?" scenarios where community members or moderators intervene. To power social simulacra, we contribute techniques for prompting a large language model to generate thousands of distinct community members and their social interactions with each other; these techniques are enabled by the observation that large language models' training data already includes a wide variety of positive and negative behavior on social media platforms. In evaluations, we show that participants are often unable to distinguish social simulacra from actual community behavior and that social computing designers successfully refine their social computing designs when using social simulacra.

CCS CONCEPTS

Human-centered computing

Collaborative and social computing systems and tools.

KEYWORDS

social computing, prototyping

ACM Reference Format:

Joon Sung Park, Lindsay Popowski, Carrie J. Cai, Menedith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2022. Social Simulaera: Creating Populated Prototypes for Social Computing Systems. In The 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22), October 29 November 2, 2022, Bend, OR, USA. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3326113.3545616

1 INTRODUCTION

How do we anticipate the interactions that will arise when a social computing system is populated [4, 23]? In social computing, design decisions such as a community's goal and rules can give rise to dramatic shifts in community norms, newcomer enculturation, and anti-social behavior [45]. Success requires that the designer make informed decisions to shape these socio-technical outcomes. Yet, despite decades of progress in research and practice, understanding the effects of these design decisions remains challenging; as a result, designers are regularly surprised by the behaviors that arise when their spaces are fully populated.

To design pro-social spaces, designers need prototyping techniques that enable them to reflect on social behaviors that may result from their design choices, then iterate [69]. Prototypes in social computing typically take the form of experience prototypes where the designer recruits a small group of people to use the system [7, 22]. However, there remains a large gap between the behaviors that arise in a small set of test users and the behaviors that arise in a socio-technical system when it is fully populated: for example, anti-social behaviors may not arise within a tight-knit group [45]; small homogeneous groups overlook the breadth of Overview: Prompting an LLM to generate thousands of users using seed personalities and generate interactions between them

Developed Solution:

Feed a community design, goals and two-line personas as input to populate the network and observe behaviours that were unexpected.

Improvements/Inferences:

- Creating more rounded user personas to make them more realistic
- Harness ability of LLM to generate the behaviours

Overview: Using data mining for predicting the performance metrics of posts published in a brands' Facebook page

Developed Solution:

Trained an SVM classifier to predict several features of the impact of a post using features from the history of posting.

Inferences:

- The paper found the most relevant feature in the model to be the content of the post
- It suggested the post content and user sentiment analysis

PAPER 2

Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach

Sérgio Moro a.b., Paulo Rita a, Bernardo Vala 1.1

- * Business Research Unit, ISCIE-University Institute of Lisbon, Portugui
- 8 ALGORITMI Research Centre, University of Minho, Portugal
- * ISCTE Business School, ISCTE-University Institute of Lisbon, Portugui

ARTICLE INFO

Article history: Received 29 September 2015 Received in revised form 11 February 2016 Accepted 15 February 2016 Available online xxxx

Keywords: Social networks Social media Data mining Knowledge extraction Sensitivity analysis Brand building

ABSTRACT

This study presents a research approach using data mining for predicting the performance metrics of posts published in brands' Facebook pages. Twelve posts' performance metrics extracted from a cosmetic company's page including 790 publications were modeled, with the two best results achieving a mean absolute percentage error of around 27%. One of them, the "Lifetime Post Consumers" model, was assessed using sensitivity analysis to understand how each of the seven input features influenced it (category, page total likes, type, month, hour, weekday, paid). The type of content was considered the most relevant feature for the model, with a relevance of 36%. A status post captures around twice the attention of the remaining three types (link, photo, video). We have drawn a decision process flow from the "Lifetime Post Consumers" model, which by complementing the sensitivity analysis information may be used to support manager's decisions on whether to publish a post.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

The worldwide dissemination of social media was triggered by the exponential growth of Internet users, leading to a completely new environment for customers to exchange ideas and feedback about products and services (Kaplan and Haenlein, 2010). According to Statista Dossier (2014), the number of social network users will increase from 0.97 billion to 2.44 billion users in 2018, predicting an increase around 300% in 8 years. Considering its rapid development, social media may become the most important media channel for brands to reach their clients in the near future (Mangold & Faulds, 2009; Korschun and Du, 2013).

Companies soon realized the potential of using Internet-based social networks to influence customers, incorporating social media marketing communication in their strategies for leveraging their businesses. Measuring the impact of advertisement is an important issue to be included in a global social media strategy (Lariscy et al., 2009). Several studies focused on finding the relationships between online publications on social networks and the impact of such publications measured by users' interactions (e.g., Cvijikj et al., 2011). However, fewer studies devoted attention to research for implementing predictive systems

 Corresponding author at; Business Research Unit, ISCTE-University Institute of Lisbon, Lisbon, Portugal, Tel.: +351 21 790 30 24, +351 25 351 03 09.

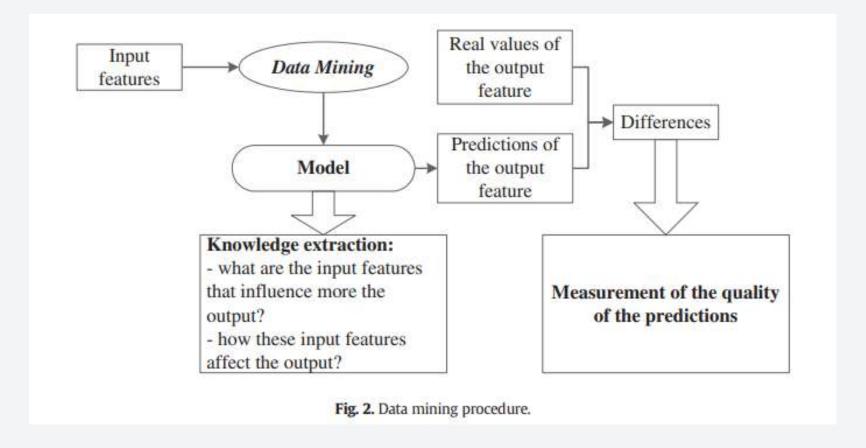
that can effectively be used to predict the evolution of a post prior to its publication. A system able to predict the impact of individual published posts can provide a valuable advantage when deciding to communicate through social media, tailoring the promotion of products and services. Advertising managers could make judged decisions on the receptiveness of the posts published, thus aligning strategies toward optimizing the impact of posts, benefiting from the predictions made. Also, it has been shown that social media publications are highly related to brand building (Edosomwan et al., 2011). Therefore, the predictive tool outlined in this paper could leverage managerial decisions to im-

Data mining provides an interesting approach for extracting predictive knowledge from raw data (Turban et al., 2011). Its application to social media has been studied, especially for evaluating market trends from users' inputs (e.g., Trainor et al., 2014). However, most of the studies focused on a reactive evaluation of what users are saying through the network, with an emphasis on gathering information from different network groups or even personal posts (e.g., Bianchi and Andrews, 2015). We focused on predicting the impact of publishing individual posts on a social media network company's page. The impact is measured through several available metrics related to customer visualizations and interactions. The predictive knowledge found enables to support manager's decisions on whether to publish each post.

For validating the taken procedure, we addressed a worldwide cos-

Features

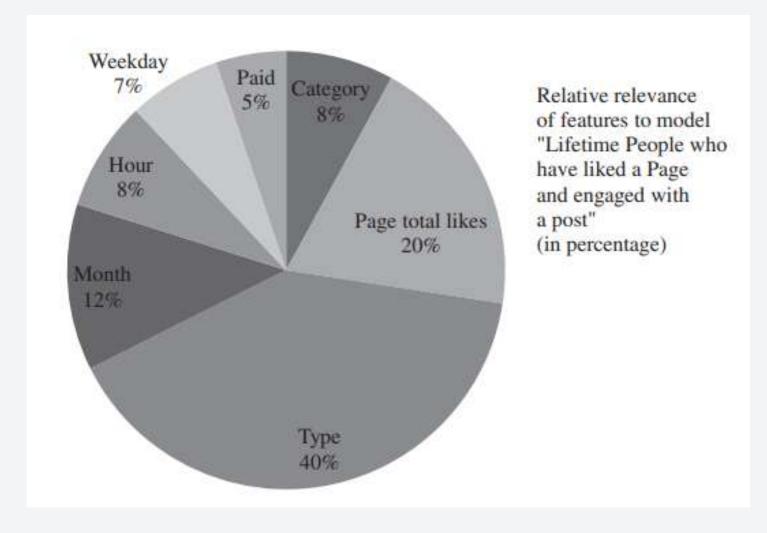
Feature	Description
Category	Manual content characterization: action (special offers and contests), product (direct advertisement, explicit brand content) and inspiration (non-explicit brand related content).
Page total likes	Number of people who have liked the company's page.
Type	Type of content (Link, Photo, Status, Video).
Post month	Month the post was published (January, February, March,, December).
Post hour	Hour the post was published (0, 1, 2, 3, 4,, 23).
Post weekday	Weekday the post was published (Sunday, Monday,, Saturday).
Paid	If the company paid to Facebook for advertising (yes, no).



Performance Metrics

Table 5
Results for performance metrics predictions

Performance metric	Mean absolute percentage error	Source of metric
Lifetime people who have liked your page and engaged with your post	26.9	
Lifetime post consumers	27.2	
Lifetime engaged users	28.8	
Lifetime post consumptions	33.1	Interactions
Shares	35.8	
Lifetime post reach by people who like your page	37.5	Visualizations
Likes	41.2	Interactions
Lifetime post impressions by people who have liked your page	47.8	Ve are
Lifetime post total reach	49.6	Visualizations
Comments	63.9	Interactions
Lifetime post total impressions	69.3	Visualizations



PAPER 3

The Impact of Content, Context, and Creator on User Engagement in Social Media Marketing

Roope Jaakonmäki University of Liechtenstein roope jaakonmaeki@uni.li Oliver Müller IT University of Copenhagen olmv@itu.dk Jan vom Brocke University of Liechtenstein jan.vom.brocke@uni.li

Abstract

Social media has become an important tool in establishing relationships between companies and customers. However, creating effective content for social media marketing campaigns is a challenge, as companies have difficulty understanding what drives user engagement. One approach to addressing this challenge is to use analytics on user-generated social media content to understand the relationship between content features and user engagement. In this paper we report on a quantitative study that applies machine learning algorithms to extract textual and visual content features from Instagram posts, along with creator- and context-related variables, and to statistically model their influence on user engagement. Our findings can guide marketing and social media professionals in creating engaging content that communicates more effectively with their audiences.

1. Introduction

Over the past decade, social media has become a popular channel through which to strengthen customers' relationships with products, brands, and companies[20], [22], [27]. In a recent survey of 3,700 marketers, 96 percent of respondents answered that they use social media for marketing [36].

However, as the number of end users and marketers who are active on social media increases [35], it becomes increasingly difficult for companies to stand out from the crowd enough to engage their target audiences. In fact, 91 percent of marketers struggle to rate at which users engage with their posts. The engagement rate measures the quantity of responses and interactions that content on social media generates from users [4], [17], [31], [38]. How the engagement rate is calculated varies across social media platforms, but it generally measures the percentage of people who react to a post in some way, such as by "liking" it or commenting on it.

The factors that drive social media engagement can be divided broadly into three groups: those that are related to the post's creator (e.g., the creator's sex, age, number of followers) [24], [21]; the post's context (e.g., time, location) [16], [41]; and certain features of the content, such as, textual content (e.g., words, tags), visual content (e.g., images, videos), and audio content. While researchers have applied various methods to study how users engage with textual content [2], [6], [8], [9], [15], [21], [24], [26], [28], [34], [38], only a few have focused on posts' visual content [4], [5], [24].

Against this background, we follow a holistic approach to study engagement in social media marketing by statistically modeling the influence on user engagement of the textual and visual features of content on user engagement while controlling for features related to creator and context. We use machine-learning algorithms to extract the textual and visual features of content from a dataset of more than 13,000 Instagram posts from professional bloggers and to identify the most important features with regards to user engagement. To the best of our knowledge, our study is among the first to use a data-analytic approach to identify automatically the most significant features that drive social media engagement. Our results can help social media marketers and users understand the most effective approach to engaging social media **Overview:** To understand the relationship between content features and user engagement

Developed Solution:

Extract textual and visual content features from Instagram posts, along with creator- and context-related variables, and to statistically model their influence on user engagement.

Table 4. Most influential content-related predictors (i.e., words, emojis, and image classes)

Word	Freq.	Coef.
Instagram	136	1 011
Switzerland	128	829
Wonderful	101	448
Video	151	393
Delicious	211	333
Sunday	309	248
Make	164	215
Blonde	121	215
Outfitoftheday	243	213
Christmas	106	194
Emoji	Freq.	Coef.
6	205	807
	301	230
类	129	204
X	190	199
65	797	185
	189	173
49	311	153
(1)	470	144
*	141	130
R	183	126

Improvements/Inferences:

- For content variables, analyzing emojis was also important
- We aim to incorporate the context of the usage of words and symbols and not just treat them as independent variables for more accurate predictions
- The importance of various words and emojis can change over time

DATA DATA DATA

DATA COLLECTION

- We needed curated data specific to a users activity over time to build a user persona
- Custom twitter scraper to scrape
 - Latest user tweets
 - Demographics (Gender, Location)
 - User stats (followers, likes, etc)
 - Liked Tweets (by the users)
 - Replies to tweets



SPECIFICATIONS

Twitter Al Information		ter API	Custom Scraper		
iniormation	Limit	Rate	Limit	Rate	
User Tweets	Hard Limit: <i>Up to</i> 10k tweets	5 calls / 15 mins	Unlimited tweets / user	50 calls / 15 minutes 1 call = 20 tweets	
User Information (Location, Gender, Following, Followers)	Unlimited	500 calls / 24 hours	Unlimited	Unlimited	
User Likes	Hard Limit: <i>Up to</i> 10k tweets	5 calls / 15 mins 200 calls / 24 hours	Unlimited	500 calls / 15 minutes 1 call = 20 liked tweets	
User Replies	No API	No API	Unlimited	50 calls / 15 minutes 1 call = 15 replies	

SAMPLE DATA

index	user_id	created	descrip	favourit	follower	followin	location	media	name	is_blue	Topics	Persona
Y	Y	7	7	Y	Y	Y	Y	Y	Y	Y	abla	7
0	15970050	Sun Aug 2	inspire so	53	1002671	43	New York	424	Simon Sinek	0	leadership,	INTJ
1	18867596	Sun Jan 11	SVP & MD,	14594	14778	1988	Germany /	1211	Sindhu Ga	0	bangalore	ENTP
2	24700996	Mon Mar 1	#Model #A	11726	112669	2236	New York,	435	Dominique	0	revolving	INTP
3	34197952	Wed Apr 2	Chairman	37	10834188	327	Mumbai, In	3938	anand mah	1	indiahope,	ENTP
4	68164273	Sun Aug 2	President,	49305	119985	1383	New Delhi	2298	Samir Saran	1	indiarespo	ENTJ
5	472816440	Tue Jan 24	DEPUTY C	75797	1085	4037	HYDERAB	7243	Swarna ku	0	nightbeaut	INTJ
6	15278055	Tue Jun 18	Early stag	6098	6306	734	Bangalore,	244	Alok Goyal	1	saas,0.496	INTJ
7	87476972	Tue Jun 13	You'll find	12892	7061	2022	Bangalore,	1028	Dravisha	1	bangalore	INTJ
8	12434706	Fri Mar 27	Hyros. Ne	5695	867730	23	Texas	2442	Alex Becke	1	invest,0.44	INTJ
9	12869043	Sat Jul 25	Security A	1862	3191	254	Pune, India	105	Rohan_lew	0	cybersecu	INFP
10	15023568	Fri Mar 11	쁓 building	162	3727	149		28	Sarah Chie	1	socialite,0	INFP

ETHICAL CONCERNS

User Privacy

- Users may not have consented to data scraping.
- Data ownership and usage ethics.

Disclosure

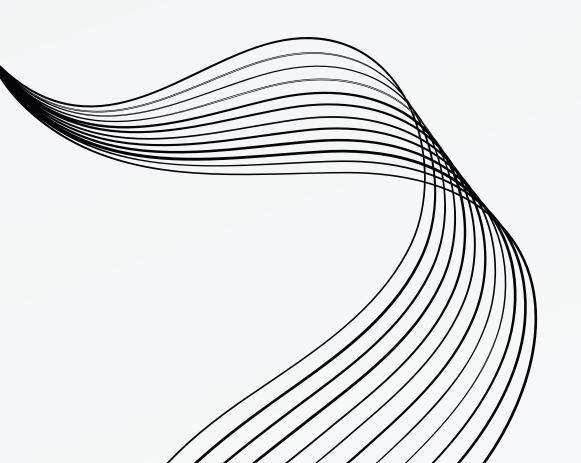
 Simulations could inadvertently reveal sensitive or personal information which they may not want public.

Exploitation

- Simulating user behaviour can be misused to exploit users.
- Making them take actions which are influenced by us.

DATA PREPROCESSING

- 1. Emoji Conversion
- 2. Handling @Mentions
- 3. Accented Word Standardization
- 4. URL Removal



- 5. Contractions Expansion
- 6.Hashtag Removal
- 7. Lemmatization
- 8. Stopword Removal

DATA PREPROCESSING

```
BACK to the Future...@elonmusk <a href="https://t.co/csuzuF6m4t">https://t.co/csuzuF6m4t</a>
How is this even possible?? Clearly she's a talented artist. But to paint 15 portraits at once i...
If you ever were wondering why such a fuss over Independence Day, just ask these two people. The...
As a child, my mother would tell me to finish all my food if I wanted to be as strong as Hanuman...
In 1947 on the cusp of Indian Independence, Winston Churchill supposedly said "...all Indian leade...
```



back future ...
even possible ? ? clearly talented artist . paint 15 portrait artits miracle ! anyone located ne...
ever wondering fuss independence day , ask two people . explain better lecture . jai hind . : in...
child , mother would tell finish food wanted strong hanumanji . later , started working gym , wo...
1947 cusp indian independence , winston churchill supposedly said ... indian leader low calibre ...

FEATURE EXTRACTION

Behavior, including personality, and sentiment per user to see how they react with the community

BEHAVIOURAL TRAITS

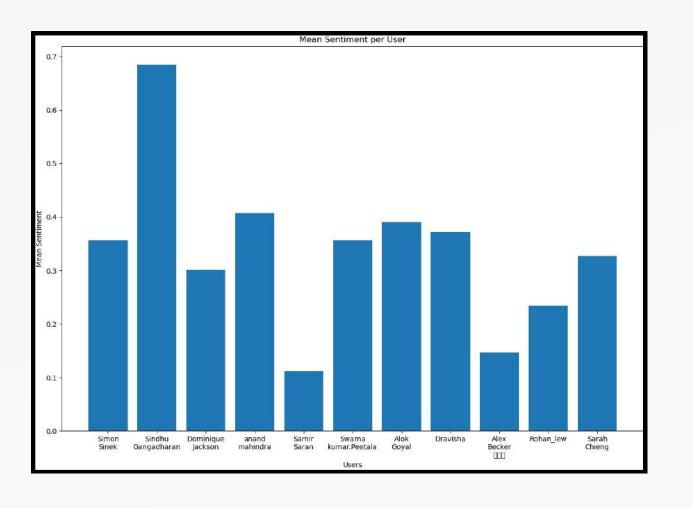
Uncovers the underlying themes and interests within a user's content regarding their preferences and affinities,.

TOPIC MODELING

BEHAVIOURAL TRAITS

Sentiment Analysis

- Utilized NLTK for sentiment analysis on tweets.
- Compiled sentiment scores for each user.
- Determined users' average sentiment disposition.



TOPIC MODELING

[(gamblingtwitter, 0.005325193753236907), (free, 0.00614130380787976), (bet, 0.00865724960473514)]

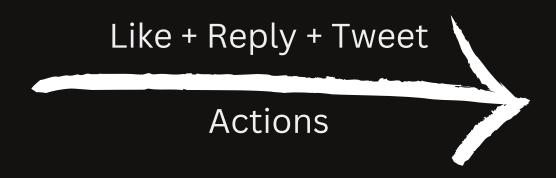
- Utilizing the YAKE extractor to derive keywords, we calculate affinity scores for each user toward various topics.
- Affinity scores are generated to gauge user preferences for specific topics.

ML METHODOLOGY



METHODOLOGY

Trained Classifier



LLM

Changes pushed to platform & starts again

Final Response



ALGORITHMS

LSTM

Classifier to predict whether a user likes/replies to a tweet

Random Forest

Classifier to generate appropriate action

Embedding Models

Convert text to trainable vectors for models

Reinforcement Lerning

Later stage network optimization

WHY LSTM?

- Language is inherently sequential, with the meaning of a word or phrase often depending on the context of surrounding words.
- LSTMs capture context through memory cells, forget gates, and input/output gates, allowing them to selectively remember and utilize information over long sequences and can capture semantic relationship with word embeddings.

LSTM MODEL

Input

Three sequential inputs (length 200 each) for 3 features, topics description and the tweet itself

One numerical input. (Sentiment_Diff)

Layers:

Embeddings: Convert sequences to 16-dimensional vectors.

LSTMs: Capture sequential patterns (64 units each).

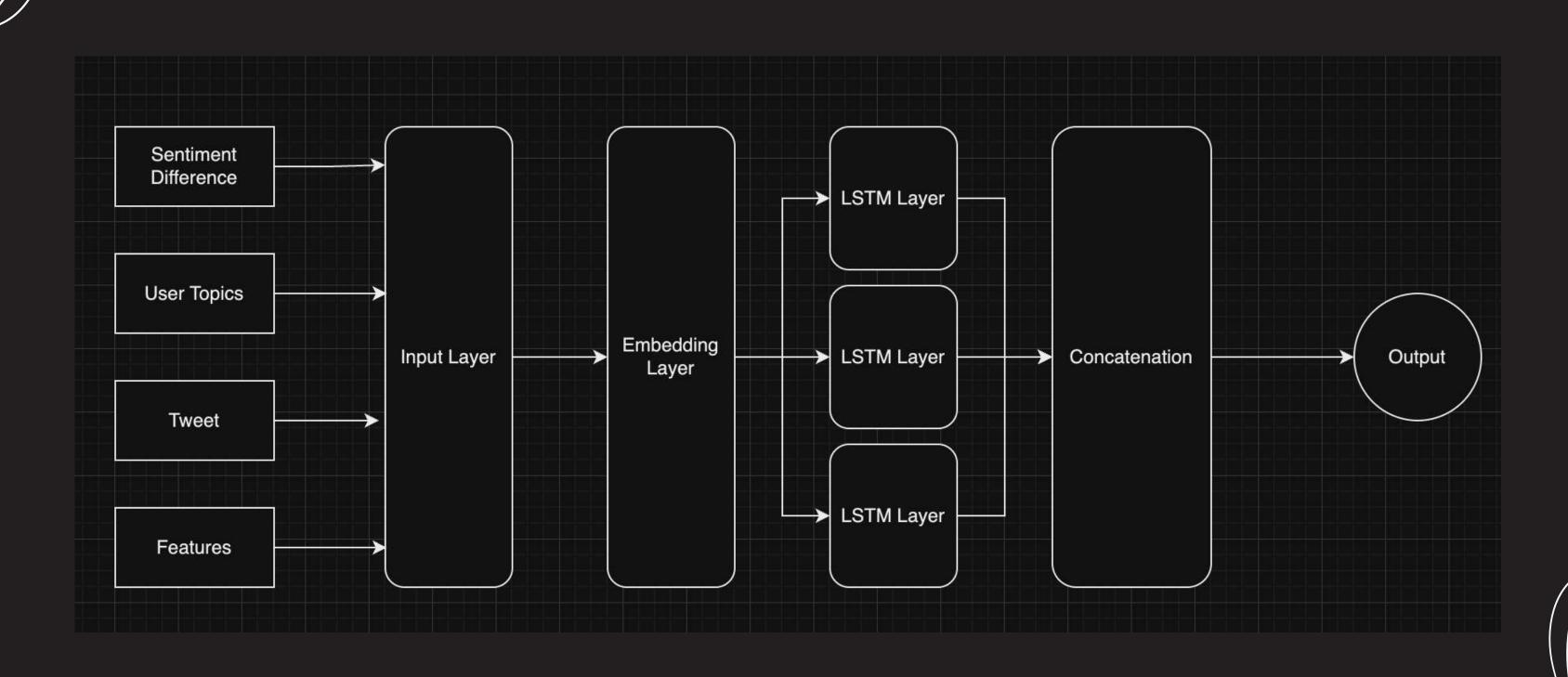
Concatenation: Merge LSTM outputs with numerical input.

Dense: Single-unit layer with sigmoid for binary output.

Parameters:

Total: 542,402

LSTM MODEL



WHY RANDOM FOREST?

- **High-Dimensional Data:** Tweets often involve high-dimensional data, where each word or n-gram can be treated as a feature.
- Sparse Feature Space: The feature space in text data is often sparse, as each tweet may only contain a subset of the available words in the entire vocabulary.
- Ensemble of Diverse Models: Tweets can vary widely in terms of language, style, and content. The ensemble nature of Random Forest, built on diverse decision trees, helps capture different patterns and relationships within the data. This diversity contributes to the model's robustness.

RANDOM FOREST MODEL

Input

- Text inputs: topics, description and the tweet itself
- Numerical input: Sentiment_Diff

Preprocessing

- Text features were tokenized and padded to be of same length
- Numerical feature was standardized

Class Weights

Class 1 indicating liked/replied to a tweet was assigned a weight of 4 due to its sparseness in the data

PERFORMANCE METRICS & DEPOLYABILITY

PREDICT USER LIKE ENGAGEMENT

Classification	Report: precision	recall	f1-score	support
0 1	0.93 0.63	0.90 0.70	0.91 0.67	2017 494
accuracy macro avg weighted avg	0.78 0.87	0.80 0.86	0.86 0.79 0.86	2511 2511 2511

Classification	Report:			
	precision	recall	f1-score	support
0	0.92	0.99	0.96	2017
1	0.95	0.66	0.78	494
accuracy			0.93	2511
macro avg	0.94	0.83	0.87	2511
weighted avg	0.93	0.93	0.92	2511

LSTM

Random Forest

PREDICT USER REPLIES ENGAGEMENT

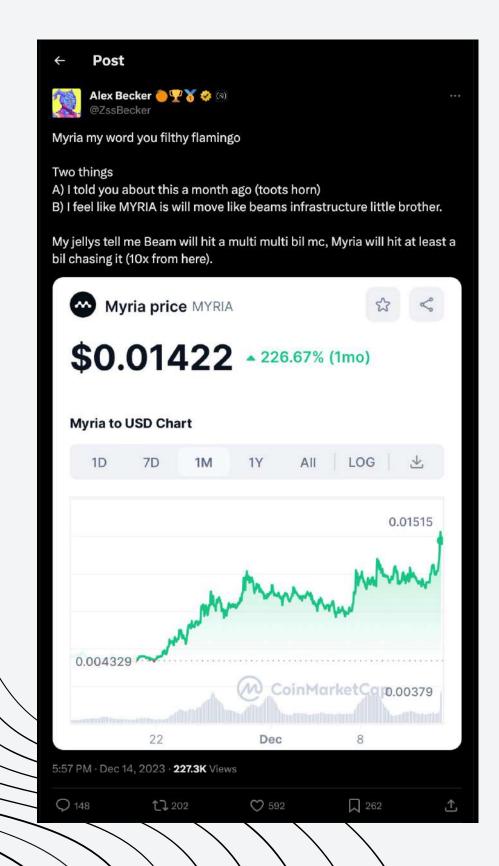
Classification	Report: precision	recall	f1-score	support
0 1	0.89 0.78	0.67 0.93	0.76 0.85	60 75
accuracy macro avg weighted avg	0.83 0.83	0.80 0.81	0.81 0.81 0.81	135 135 135

Classification	Report: precision	recall	f1-score	support
0	0.89	0.82	0.85	60
1	0.86	0.92	0.89	75
accuracy			0.87	135
macro avg	0.88	0.87	0.87	135
weighted avg	0.88	0.87	0.87	135

LSTM

Random Forest

REAL WORLD TEST

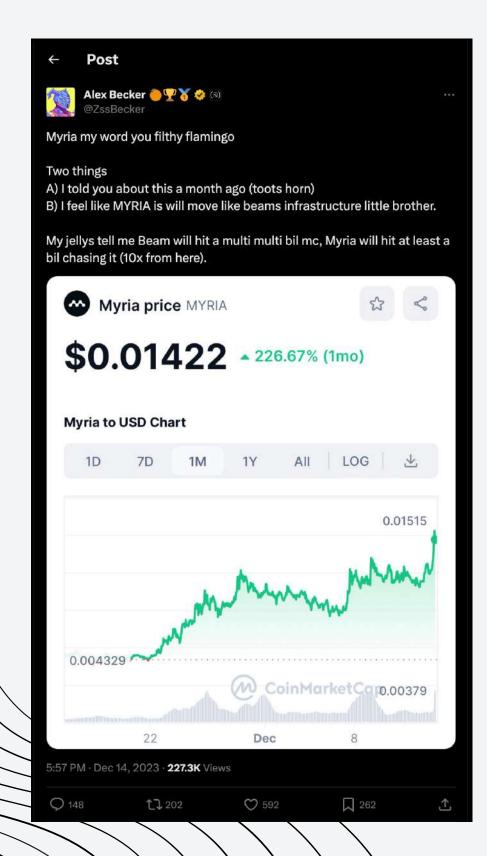






Test run with unseen users on an unseen tweet for a real world test.

REAL WORLD TEST



	Predicted	Actual	
Random Forest	624 Likes	634 Likes	
Random Forest	121 Comments	148 Comments	
LCTM	639 Likes	634 Likes	
LSTM	113 Comments	148 Comments	



 OpenAl APIs were used to generate comments on the users behalf.

"Myria my word you filthy flamingo

Two things

- A) I told you about this a month ago (toots horn)
- B) I feel like MYRIA is will move like beams infrastructure little brother.

My jellys tell me Beam will hit a multi multi bil mc, Myria will hit at least a bil chasing it (10x from here)."

It seems like you are quite passionate about the potential of Myria and Beam. It's always exciting to see the potential for growth in the market. Can you share more about what you find promising about these projects?









CHALLENGES

Network Size

We will need sufficient network size to properly simulate and assess the network effect

Computation

As the network size grows, more computational power would be required to simulate and maintain individual users

Generalization

A representative network needs to be constructed to accurately represent a real world scenario

