

Mid Prep



Adobe

BEHAVIOUR SIMULATION CHALLENGE



About Adobe Digital Experience Business

Adobe Experience Cloud gives digital businesses everything they need to design and deliver great customer experiences. Adobe solutions integrate best-in-class Adobe products to help its customers tackle pressing business challenges.

1. **Experience-led growth:** Adobe Experience Cloud gives our business clients the insights and tools to create individual and engaging experiences that their customers are looking for — the ones that drive growth for our customers' B2B or B2C business.
2. **Content supply chain:** Adobe's customers can build a scalable and efficient content supply chain with the help of Adobe GenStudio. This powerful solution combines best of content ideation, creation, production, and activation with the powers of generative AI.
3. **Personalization at scale:** Adobe Experience Cloud lets our clients reach every customer with engaging, personalized experiences — right when they need it. Our technology gives them real-time unified data and insights, connected customer journeys, and AI-assisted workflows for one-to-one moments and planned campaigns.
4. **B2B marketing:** Experience Cloud lets our customers deliver account-based experiences that differentiate their brand. Advanced features such as real-time data and account profiles and AI-driven workflows make every customer interaction relevant and let them deliver a seamless digital journey that's directly integrated with the sales process.

Background

Timely and targeted content generation is essential to yield desired marketing results for any business. This is carried through higher user engagement which in turn drives sales and broader customer outreach. Scoring high on user engagement KPIs boosts brand credibility, reach and product recall.

This also helps marketers strengthen on *algorithmic favorability* which for a social media company decides on the post that gets propelled based on the engagement it drives.



Problem Description

The process of communication is defined by marketing researchers as below:

A receiver, upon receiving a message from a sender over a channel, interacts with the message, thereby generating *effects* (*user behavior*). Any message is created to serve an end goal. For a marketer, the eventual goal is to get the desired effect (user behavior) i.e. such as likes, comments, shares and purchases, etc.



In this challenge, you will try to solve the problem of behavior simulation (Task-1) and content simulation (Task-2), thereby helping marketers to estimate user engagement on their social media content as well as create content that elicits the desired key performance indicators (KPI) from the audience.

(You can read these papers [\[1\]](#), [\[2\]](#) to understand both the tasks a bit better)

Dataset

Brands use Twitter to post marketing content about their products to serve several purposes, including ongoing product campaigns, sales, offers, discounts, brand building, community engagement, etc. User engagement on Twitter is quantified by metrics like user likes, retweets, comments, mentions, follows, clicks on embedded media and links. For this challenge, we have sampled tweets posted in the last five years from Twitter enterprise accounts. Each sample contains tweet ID, company name, username, timestamp, tweet text, media links and user likes.



Problem Statement

Task-1: Behavior Simulation

- Given the content of a tweet (text, company, username, media URLs, timestamp), the task is to predict its user engagement, measured by likes.
- Sample IO

Input:

date	content	username	media	inferred company
2019-07-05 11:32:00	Sandy, muddy terrain exploration...	Toyota_Fortuner	[Photo(previewUrl='https://pbs.twimg.com/media/..').]	toyota

Output:

likes
10

- Size of train data: 300K samples
- The results will be evaluated under 2 regimes (10K samples each):
 - Predicting likes for tweets of unseen brands (test dataset) but seen time period (training dataset: Jan-2018 to Dec-2020)
 - Predicting likes for tweets of unseen time period (test dataset) but seen brands (training dataset)

Link to [Training Dataset](#)

NOTE: Test dataset for both tasks will be shared 2 days prior to final submission.



Task 2: Content Simulation

- Given the tweet metadata (company, username, media URL, timestamp), generate the tweet text.
- Size of train data: 300K samples
- Sample IO

Input:

date	likes	username	media	inferred company
2019-07-05 11:32:00	10	Toyota_Fortuner	[Photo(previewUrl='https://pbs.twimg.com/media/..)]	toyota

Output:

content
Sandy, muddy terrain exploration? Yes – we’ve got you covered. Visit ...

- The results will be evaluated under 2 regimes (10K samples each):
 - Predicting tweet text for tweets of unseen brands (test dataset) but seen time period (training dataset: Jan-2018 to Dec-2020)
 - Predicting tweet text for tweets of unseen time period (test dataset) but seen brands (training dataset)

Link to [Training Dataset](#)

NOTE: Test dataset for both tasks will be shared 2 days prior to final submission.



Evaluation Guidelines

Submission Deadline - **14th December**

Testing Dataset will be shared on 13th December

Evaluation metric(s) – **50% weightage**

For Task 1:

1. RMSE (Root Mean Squared Error) between predicted and ground-truth likes.

For Task 2:

1. BLEU 1-4
2. ROUGE
3. CIDER

Refer this [\[1\]](#) for the above metrics.

The metrics will be used to compare predicted and ground-truth tweet text.

Approach – **35% weightage**

1. Efficiency of solution
2. Novelty of approach

Presentation – **15% weightage**

Submission Guidelines

1. Link to GitHub repository (access made public on the day of final submission)
2. Submit a report detailing your approach, results, or any inference/assumption you feel is important. The maximum length can be 3 pages in ACL LaTeX format (excluding appendix and references).

Dataset shared is open source. The IP to final solution will belong with Adobe.

