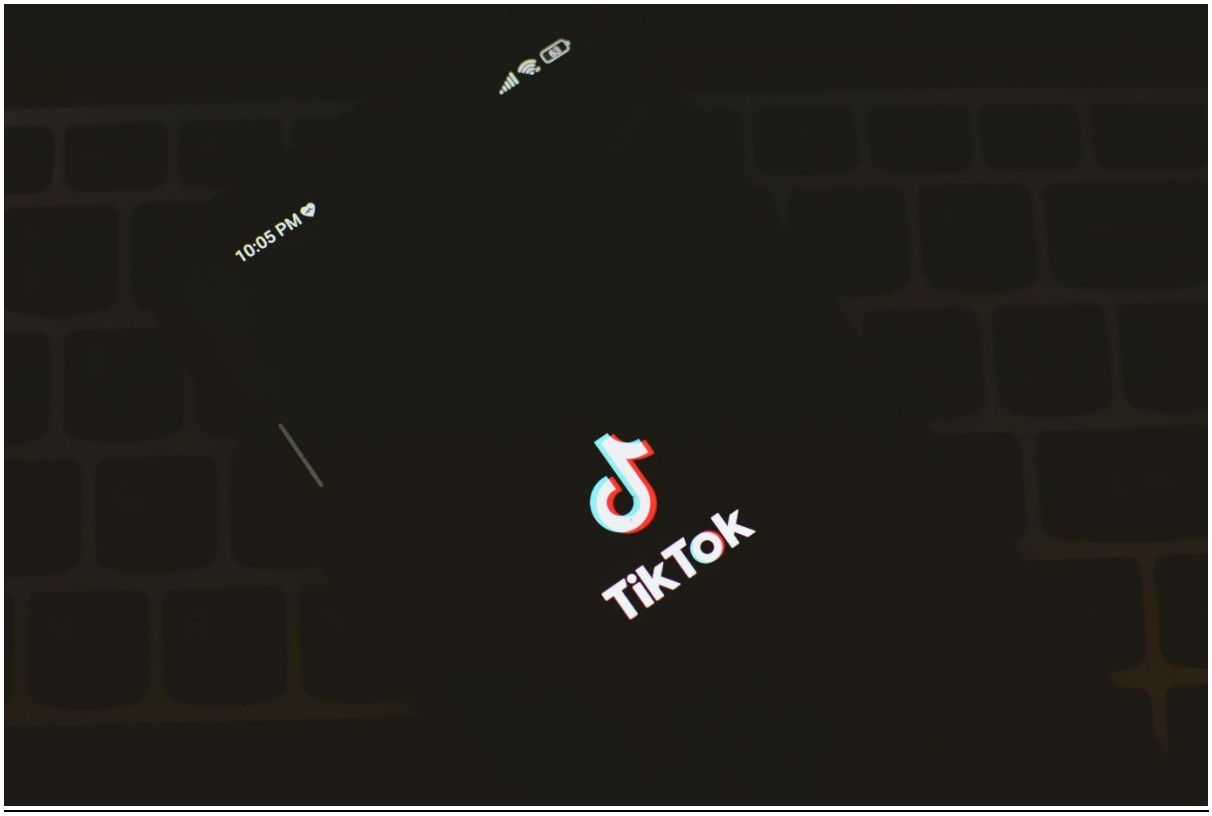
TikTok Machine Learning Model Project



Note: *This project is created in partnership between Google Advanced Data Analytics Professional Certificate and the short-form video hosting company, TikTok. The story, all names, characters, and incidents portrayed in this project are fictitious. No identification with actual persons (living or deceased) is intended or should be inferred. And the data shared in this project has been created for pedagogical purposes.*

Project background

TikTok users have the ability to report videos and comments that contain user claims. These reports identify content that needs to be reviewed by moderators. This process generates a large number of user reports that are difficult to address quickly.

TikTok is working on the development of a predictive model that can determine whether a video contains a claim or offers an opinion. With a successful prediction model, TikTok can reduce the backlog of user reports and prioritize them more efficiently.

Scenario 1

As a data analyst on TikTok's data team, with new considerations from the leadership team, my responsibility is to create a project proposal by assigning the required data analytical tasks into realistic milestones that will advise future steps in the claims classification project.

For my first assignment, I will be providing TikTok with a project proposal that will create milestones for the tasks within the comment classification project.

Team members at TikTok

As a new data analyst, I will be working closely with a talented team of experienced data professionals. I'll also collaborate with TikTok co-workers outside the data team as the project requires.

Data team roles

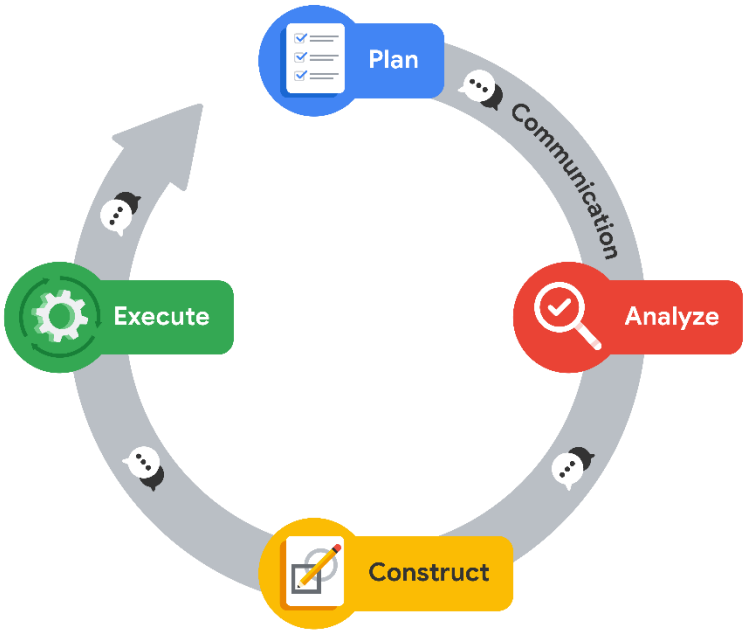
- Willow Jaffey- Data Science Lead
- Rosie Mae Bradshaw- Data Science Manager
- Orion Rainier- Data Scientist

Cross-functional team members

- Mary Joanna Rodgers- Project Management Officer
- Margery Adebowale- Finance Lead, Americas
- Maika Abadi- Operations Lead

Workflow Structure

I will be using the PACE workflow framework throughout this project in order to provide a clear foundation and structure for this data analysis project. PACE is an acronym and each one of the letters represents an actionable stage in a project: plan, analyse, construct, and execute.



Plan Stage: First, I need to ask and answer some relevant questions for this project. These include:

- Who is my audience for this project? The answer to this question is necessary as it would help me identify my stakeholders for the project. My audience for this project includes my data team members, the cross-functional team members, the TikTok users who may report a claim or opinion, and the moderators who review the videos.
- What am I trying to solve or accomplish? And what do I anticipate the impact of this work will be on the larger needs of the client? These would help me identify and be familiar with the goal of the project. I am trying to develop a predictive model capable of distinguishing whether a

video contains a claim or an opinion. This aims to address the issue of the time-consuming process moderators face when evaluating each video, user comment, and content claim. By building a machine learning model that can accurately classify user interaction data, we aim to streamline and expedite the moderators' work. The anticipated impact is an improved experience for TikTok users, ensuring their claim submissions are handled promptly.

- What questions need to be asked or answered? These include:
    - What type of data will we be working with on this project?
    - What is the condition of the provided dataset?
    - What variables will be the most useful?
    - Are there trends within the data that can provide insight?
    - What steps can I take to reduce the impact of bias?
    - Which team member would be assigned to which tasks?
    - What information will be the most useful in TikTok's data?
    - What type of regression model should we use for this project?
    - What method of statistical testing should be used for the project?
    - What is the best method of hypothesis testing for this dataset?
- What resources are required to complete this project?
    - Team members
    - Input from stakeholders
    - Budget/Funding
    - Project Dataset
    - A statistical tool – e.g., Python (Python notebook)
- What are the deliverables that will need to be created over the course of this project?
    - A project proposal
    - Prepared and cleaned datasets
    - Statistical model
    - Regression analysis model
    - Machine learning model
    - Exploratory Data Analysis reports
    - Stakeholder reports
    - Visualizations e.g. dashboard