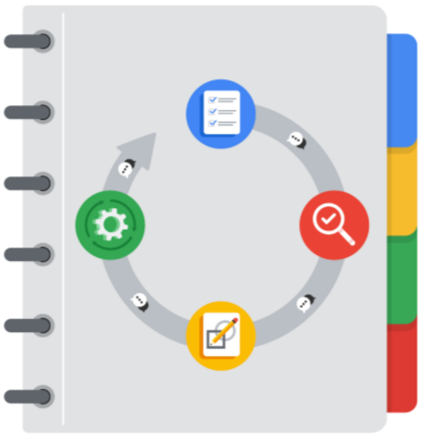
# TikTok PACE Strategy Document VI

*Machine Learning Model*

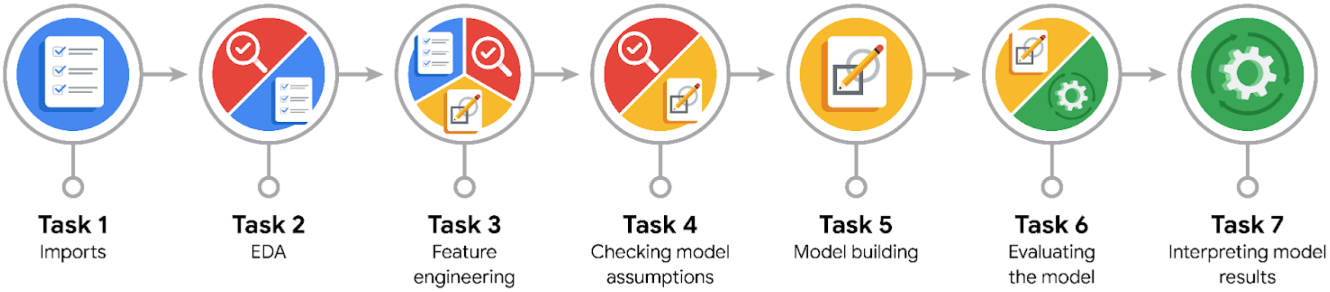
## Introduction



PACE stands for Plan, Analyze, Construct and Execute. It is a framework that illustrates the foundation and structure for data analysis projects and each letter represents an actionable stage in a project. The stage "Plan" involves the definition of the project scope, the research of business data and the workflow development. The stage "Analyze" involves data scrubbing, data conversion and database formatting. The stage "Construct" involves building models and machine learning algorithms and selecting a modeling approach. The stage "Execute" involves the presentation of results to decision-makers, stakeholders and others in order to receive feedback. This framework is built upon an iterative cycle where each stage may reveal new insights, requiring the return to earlier stages. A PACE strategy document is used to record decisions and reflections at different stages of the data analytical process. It typically includes the definitions of roles and actions to ensure clarity and accountability.

## Purpose

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. For this stage of the project, we identify seven main tasks that are presented in the following visual.



## Considerations

### PACE: Plan Stage

- What problem are we trying to solve or what outcome are we trying to achieve?

We aim to classify TikTok videos as either claims or opinions to prioritize potentially harmful content for review.

- Who are the external stakeholders for this project?

The external stakeholders include TikTok's content moderation team and possibly departments involved in managing user interactions and claim status.

- What resources would be useful for this stage?

Historical engagement data, labeled claim/opinion video data and machine learning tools like Python libraries for modeling.

- Are there any ethical issues to consider at this stage of the project?

The risk of false negatives is critical, as it could allow videos that violate terms of service to remain unchecked.

- Is the data accurate and reliable?

The data is accurate and balanced, with minimal missing values, making it suitable for modeling.

- What data would be perfect for answering our research question?

Detailed engagement metrics along with precise labels for videos as claims or opinions would be ideal.

- What data do we have access to or can we obtain?

We have access to video engagement data, video content and labels indicating whether each video is a claim or an opinion.

- What metric should be used to measure the success of our business/organizational objective?

Recall should be used, as it ensures that most claim videos are identified, minimizing the risk of harmful content slipping through.

## PACE: Analyze Stage

- Based on this stage, has our research question changed? Does the plan need to be adjusted?

The research question remains the same and the plan is focused on optimizing model performance to identify claims.

- Does the data violate the assumptions of the model? Is this a significant issue?

The data does not violate key assumptions for tree-based models, as they handle outliers and imbalanced features well.

- Why did we choose the specific independent variables for our model?

Engagement metrics like views, likes, shares and downloads were chosen because they are strong predictors of claim status.

- Why is Exploratory Data Analysis (EDA) important before building a model?

It helps to understand the distribution of data, identify key predictors and guide feature selection.

- What has the data exploration revealed?

The analysis revealed that engagement metrics are highly indicative of whether a video is a claim.

- What resources would be useful for this stage?

Data visualization tools for EDA and statistical analysis libraries to validate assumptions and identify key predictors.

## PACE: Construct Stage

- Are there any unusual or unexpected patterns in the data? Is this a problem and can it be addressed?

No significant unexpected patterns were found that impact model performance, so no major issues need to be addressed.

- Which independent variables did we choose for the model?

Key variables include video view count, like count, share count and download count, which reflect user engagement levels. Another set of features is derived from the vectorized video transcript.

- How well does our model fit the data and what is its validation score?

The Random Forest model fits the data well, achieving a recall score of 99%.

- Can we make the model better? Are there any changes we would consider?

The model's performance is near perfect, but adding features like video report counts could offer slight improvements.

- What resources would be useful for this stage?

Computational resources for model training and tuning, as well as access to advanced machine learning libraries.

## PACE: Execute Stage

- What important discoveries have we made from our model? Can we explain our model?

The model's predictions rely heavily on video engagement metrics, which are the strongest indicators of claims.

- What are the criteria for model selection?

The model with the highest recall score was selected to prioritize detecting as many claim videos as possible.

- Does the model make sense and are the final results acceptable?

The model makes sense as it aligns with engagement-driven behavior and the results are highly accurate.

- Do we believe the model could be enhanced? If so, how?

Adding features like the number of times a video was reported could enhance the model's precision.

- Were there any features that were not important at all? What would happen if they were removed?

Less predictive features did not significantly impact the model, but removing them would not drastically alter performance due to the strong signal from engagement metrics.

- What recommendations would we provide to the organization based on the model built?

The model can be deployed with continuous monitoring of engagement trends. Additional data subsets should be evaluated for robustness.

- Given our knowledge of the data and model, what other questions could we explore for the team?

We could investigate whether specific video categories are more prone to being classified as claims.

- What resources would be useful for this stage?

Data monitoring tools to track the distribution of features post-deployment and A/B testing frameworks for continuous improvement could be valuable.

- Are there any ethical issues to consider at this stage?

The impact of false positives should be monitored to avoid unnecessary reviews of opinion videos.

- When the model makes a mistake, what is the underlying cause? How does this impact the use case?

Mistakes occur primarily due to high engagement on opinion videos, leading to false positives. This may cause non-violating videos to be reviewed unnecessarily, which could strain moderation resources.