# Executive Summary

Milestone # 06

## ISSUE / PROBLEM

TikTok team is adamant to pick out ideas and classify them as claims or opinions. The team had previously performed analysis .Now we are building the model for the predictions.

## RESPONSE

Building a machine learning model that can be used to determine whether a video contains a claim or whether it offers an opinion. With a successful prediction model, TikTok can reduce the backlog of user reports and prioritize them more efficiently. **claim_status**. is a binary value that indicates whether a video is a claim or an opinion. This will be the target variable.This is a classificatic task because the model is predicting a binary class.

## IMPACT

This step finally solve the problem statement TikTok had been eying till now. Predicting whether an idea is claim or opinion

## ANALYSIS

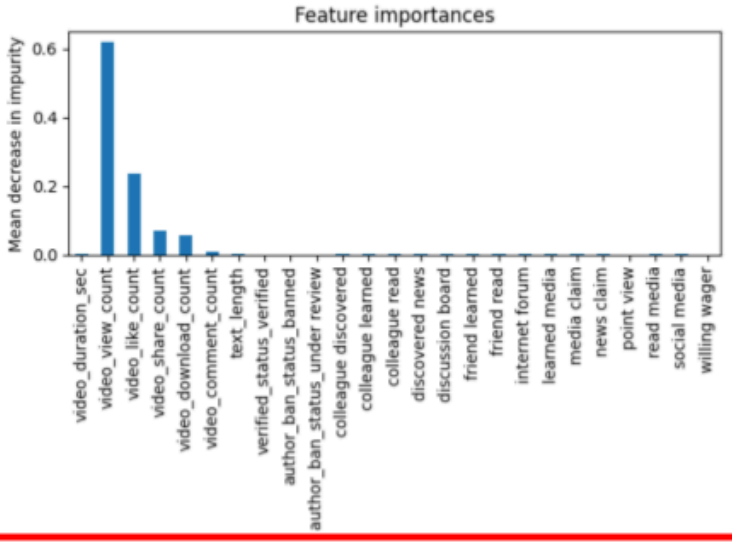
Built **Random Forest** and **XGBoost Model** Compared the finding of both on the basis of recall because it's more important to minimize the mode wrongfully predicts a video is an opinion when in fact it is a claim.Recall score of Random Forest ,

```
[46]: 0.9994806638131711
```

Recall score of XGBoost Model

```
[51]: 0.9989540885869099
```

Both are excellent predictor of target variable but Random Forest gave higher and near to perfect recall score hence it was chosen.
Trained and tested Random Forest model and determined which feature holds importance in predicting claim_status most




## KEY INSIGHTS

- This model performs well on both the validation and test holdout data. Furthermore, both precision and F1 scores were consistently high. The model very successfully classified claims and opinions.
- The model's most predictive features were all related to the user engagement levels associated with each video. It was classifying videos based on how many views, likes, shares, and downloads they received.
- The current version of the model does not need any new features. However, it would be helpful to have the number of times the video was reported. It would also be useful to have the total number of user reports for all videos posted by each author.