

Phase 6: Automated Content Moderation: Machine Learning for Optimized Classification of TikTok Claims vs. Opinions

Email from Project Management Officer

Subject: Approaching the final claims model

From: "Rodgers, Mary Joanna" —maryjoannarodgers@tiktok

Cc: "Rainier, Orion" —orionrainier@tiktok; "Jaffey, Willow" —willowjaffey@tiktok;
"Bradshaw, Rosie Mae" —rosiemaebradshaw@tiktok; "Abadi, Maika," —
maikaabadi@tiktok;

Hello TikTok team,

The leadership team would like to recognize what an excellent job the team has done on the executive summaries throughout this project. Your hard work and dedication to assisting TikTok users is greatly appreciated.

Operations Lead, Maika Abadi brought to our attention some concerns from stakeholders in other departments. As your team continues to move forward building the final predictive model, the leadership team would like you to consider the following:

1. Which modeling solution did the team choose for the final claims model?
2. What criteria did the data team use to determine their modeling approach?
3. What are the ethical implications of the final claims model?
4. What are the consequences if our claims model makes errors?
5. Do the benefits of such a model outweigh the potential problems?
6. After evaluating the model, should TikTok proceed with using this model?
7. Consider the impact that "banned authors" have on claims or opinions. To what extent do those authors make claims or opinions?

Note: For the purposes of this fictional project, "Banned authors" are users who submitted videos that "violated the terms of service."

1. With the results from your model, consider what the next steps would be for future video reviews, or downstream evaluation. Would TikTok need a team member to

review the video content? If so, at what part of the process would a review be the most useful?

Perhaps these questions can guide your final executive summary? Also, include an idea of the confidence your team has in the accuracy of the claims model.

Thank you for your great work,

Mary Joanna Rodgers

Project Management Officer

TikTok

—

Network with TikTok employees from a variety of teams and locations. Participate in TikTok Tuesdays, every Tuesday @2pm EST.

Email from the Data Science Lead

Subject: Claim model considerations

From: “Jaffey, Willow” —willowjaffey@tiktok

Cc: “Bradshaw, Rosie Mae” —rosiemaebradshaw@tiktok

Happy Monday!

You have done great work so far. We are in the final stage of the claims classification project and I am excited to see the final claims model and how it can help TikTok streamline the claims process overall.

I’ve determined that an effective final model could be approached in several ways, such as a Naive Bayes or tree-based model. However, having looked at the data analysis so far, I believe our most effective method for the sample dataset is a random forest model.

Once complete, please send an executive summary of your random forest model and results to Rosie Mae and myself so we can prepare to present the results to the leadership team. Be sure to include what was requested in the email from Mary Joanna, a summary of the variables used to make the prediction, and an idea of the accuracy of the model.

I look forward to seeing what you build!

Willow Jaffey

Data Science Lead

TikTok