Phase 5: Predictive Modeling for TikTok Verification: Multiple Logistic Regression Analysis

Email from Maika Abadi, Operations Lead

Subject: Details on Regression Model

From: "Abadi, Maika,"— maikaabadi@tiktok

Cc: "Jaffey, Willow" —willowjaffey@tiktok; "Rodgers, Mary Joanna" — maryjoannarodgers@tiktok; "Bradshaw, Rosie Mae" rosiemaebradshaw@TikTok; "Rainier, Orion"—orionrainier@tiktok

Hello Data Team,

I really appreciate your work, and thanks for the explanation of the next phase of the data analysis.

I'm curious to know more about how different variables are associated with whether a user is verified. I was hoping to get a bit more detail on the regression your team is planning to conduct. Will you be applying a linear regression model or a logistic regression model? It wasn't clear in the meeting, and I wanted to be sure our teams are aligned on expectations.

Thank you,

Maika Abadi

Operations Lead

TikTok

Email from Rosie Mae Bradshaw TikTok's Data Science Manager

Subject: RE: Details on Regression Phase

From: "Bradshaw, Rosie Mae" —rosiemaebradshaw@TikTok

Cc: "Jaffey, Willow" —willowjaffey@tiktok; "Rodgers, Mary Joanna" — maryjoannarodgers@tiktok; "Rainier, Orion"—orionrainier@tiktok; "Abadi, Maika,"— maikaabadi@tiktok

Thank you for your email.

Apologies that it was not clear in the meeting.

To answer your question, we've decided to look into how to predict 'verified_status', which we believe will help us understand how video characteristics relate to verified users. To achieve this, the data team will build a logistic regression model using 'verified_status' as the outcome variable. The results of this milestone will inform us as we approach constructing the final claims prediction model.

Feel free to reach out with additional questions.

Many thanks,

Rosie Mae Bradshaw

Data Analysis Manager

TikTok

Email from Rosie Mae Bradshaw TikTok's Data Science Manager

Subject: RE: Details on Regression Phase

From: "Bradshaw, Rosie Mae" —rosiemaebradshaw@TikTok

Cc: "Rainier, Orion"—orionrainier@tiktok

Hello my Data team!

Would you two mind completing the following using the fictional test data set::

- Logistic regression model in a Python notebook based 'verified_status' variable in the claims classification dataset.
    - Be sure to include a confusion matrix of the results and the accuracy score of the model
- Draft an executive summary of your results

I'd appreciate a chance to look it over before you send it over to Mary Joanna. Please write the summary as if you're addressing the leadership team.

Best regards,

Rosie Mae Bradshaw

Data Analysis Manager

TikTok