Scenario

The data team at TikTok are close to completing the claims classification project. Earlier, the data team completed a project proposal, and used Python to explore and analyze the claims dataset, created data visualizations, and conducted a statistical test. Most recently, the team built a logistic regression model to predict whether statements are claims or opinions within submissions, based on the 'verified_status' variable within the claims dataset.

Operations Manager, Maika Abadi mentioned that there are some questions from stakeholders in other departments, which he would like to share before the process of building the final begins.

Data Science Lead, Willow Jaffey determined that an effective final model could be approached in several ways, such as Naive Bayes or tree-based models. However, Willow and the team have decided to develop a random forest model during the evaluation stage for this project.

At the start of a new week, you receive an email from Mary Joanna Rodgers, the Project Management Officer. TikTok leadership has some consideration for the data team as the team approaches the final model. You also receive a follow-up email from Willow asking you to build the final model and prepare an executive summary to share the results.

Note: Team member names used in this workplace scenario are fictional and are not representative of TikTok.

_____

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:

Which modeling solution did the team choose for the final claims model?

What criteria did the data team use to determine their modeling approach?

What are the ethical implications of the final claims model?

What are the consequences if our claims model makes errors?

Do the benefits of such a model outweigh the potential problems?

After evaluating the model, should TikTok proceed with using this model?

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."

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

Project background

TikTok's data team is near the end of the claims classification project. The following tasks are needed to complete the project:
- Model building
- Model evaluation
- Summarize findings for cross-departmental stakeholders within TikTok

Your assignment

You will create the final machine learning model for the claims classification data. You will be responsible for leading these final tasks, which include feature engineering, model development, and evaluation.

Specific project deliverables

With this end-of-course project, you will gain valuable practice and apply your new skills as you complete the following:

- Course 6 PACE Strategy Document to consider questions, details, and action items for each stage of the project scenario
- Answer the questions in the Jupyter notebook project file
- Build a machine learning model
- Evaluate the model
- Interpret the results
- Create an executive summary to share your results with cross-departmental stakeholders within TikTok

Good luck with this project! Your team members at TikTok are looking forward to seeing how you communicate your creative work and approach problem-solving!