Activity Overview

In this activity, you will showcase your ability to use Python to build classification models. You will also update team members and stakeholders through an executive summary, demonstrating your ability to organize and communicate key information.

For additional information on how to complete this activity, review the previous readings:

End-of-course project introduction ↗ and *Course 6 end-of-course portfolio project overview: TikTok* ↗.

Be sure to complete this activity before moving on. The next course item will provide you with completed exemplars to compare to your own work. You will not be able to access the exemplars until you have completed this activity.

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

Step-By-Step Instructions

Follow the instructions to complete the activity. Then, go to the next course item to compare your work to a completed exemplar.
Step 1: Access the templates

To use the templates for this course item, click each link below and select *Use Template*.

Link to templates:

Course 6 PACE strategy document ↗

Course 6 Executive summary ↗

OR

If you do not have a Google account, you can download the templates directly from the attachments below:

🔗 Activity Template_ Course 6 PACE strategy document
DOCX File

🔗 Activity Templates_ Executive summaries
PPTX File

⟩ Step 2: Access the end-of-course project lab

Note: The following lab is also the next course item. Once you complete and submit your end-of-course project activity, return to the lab instructions' page and click Next to continue on to the exemplar reading.
To access the end-of-course project lab, click the following link and select *Open Lab*.

Course 6 TikTok project lab ↗

Your Python notebook for this project includes a guided framework that will assist you with the required coding. Input the code and answer the questions in your Python notebook to build a logistic regression model. You'll find helpful reminders for tasks like:
Ethical considerations

Feature engineering

Model building and evaluation

You will also discover questions in this Python notebook designed to help you gather the relevant information you'll need to write an executive summary for your team.
Use your completed PACE strategy document and Python notebook to help you prepare your executive summary.

⟩ Data Dictionary

This project uses a dataset called tiktok_dataset.csv. It contains synthetic data created for this project in partnership with TikTok. Examine each data variable gathered.
19,383 rows – Each row represents a different published TikTok video in which a claim/opinion has been made.
12 columns

Column name	Type	Description
#	int	TikTok assigned number for video with claim/opinion
claim_status	obj	Whether the published video has been identified as an "opinion" or a "claim." In this dataset, an "opinion" refers to an individual's or group's personal belief or thought. A "claim" refers to information that is either unsourced or from an unverified source.
video_id	int	Random identifying number assigned to video upon publication on TikTok
video_duration_sec	int	How long the published video is measured in seconds
video_transcription_text	obj	Transcribed text of the words spoken in the published video
verified_status	obj	Indicates the status of the TikTok user who published the video in terms of their verification, either "verified" or "not verified"
author_ban_status	obj	Indicates the status of the TikTok user who published the video in terms of their permissions: "active," "under scrutiny," or "banned"
video_view_count	float	The total number of times the published video has been viewed
video_like_count	float	The total number of times the published video has been liked by other users
video_share_count	float	The total number of times the published video has been shared by other users
video_download_count	float	The total number of times the published video has been downloaded by other users
video_comment_count	float	The total number of comments on the published video

⟩ Step 3: Complete your PACE strategy document

The Course 6 PACE strategy document includes questions that will help guide you through the Course 6 TikTok workplace scenario project. Answer the questions in your PACE strategy document to prepare for using Python to inspect and organize your data.

As a reminder, the PACE strategy document is designed to help you complete the contents for each of the templates provided. You may navigate back and forth between the PACE strategy document and the Python notebook. Make sure your PACE strategy document is complete before preparing your executive summary.

Step 4: Prepare an executive summary

Your executive summary will keep your teammates at TikTok informed of your progress. The one-page format is designed to respect teammates and stakeholders who may not have time to read and understand an entire report.

First, select one of the executive summary design layouts from the provided template. Then, add the relevant information. Your executive summary should include the following:

A summary of the benefits and limitations of your model

The results of your analysis

Recommendations or insights based on your results

Complete your executive summary to effectively communicate your results to external stakeholders.

Pro Tip: Save the templates

Finally, be sure to save a blank copy of the templates you used to complete this activity. You can use them for further practice or in your professional projects. These templates will help you work through your thought processes and demonstrate your experience to potential employers.

What to Include in Your Response

Later, you will have the opportunity to assess your performance using the criteria listed below. Be sure to address the following elements in your completed activity.

Course 6 PACE strategy document:

Answer the questions in the PACE strategy document

Course 6 Python notebook:

Build classification models

Course 6 Executive summary:

Clearly articulate the challenges presented in this data project

Identify the outcome of your work

Include recommendations for future work/next steps