## Course Six

# The Nuts and Bolts of Machine Learning

### Instructions

Use this PACE strategy document to record decisions and reflections as you work through the end-of-course project. As a reminder, this document is a resource that you can reference in the future and a guide to help consider responses and reflections posed at various points throughout projects.

### Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

- [ ] Complete the questions in the Course 6 PACE strategy document
- [ ] Answer the questions in the Jupyter notebook project file
- [ ] Build a machine learning model
- [ ] Create an executive summary for team members and other stakeholders

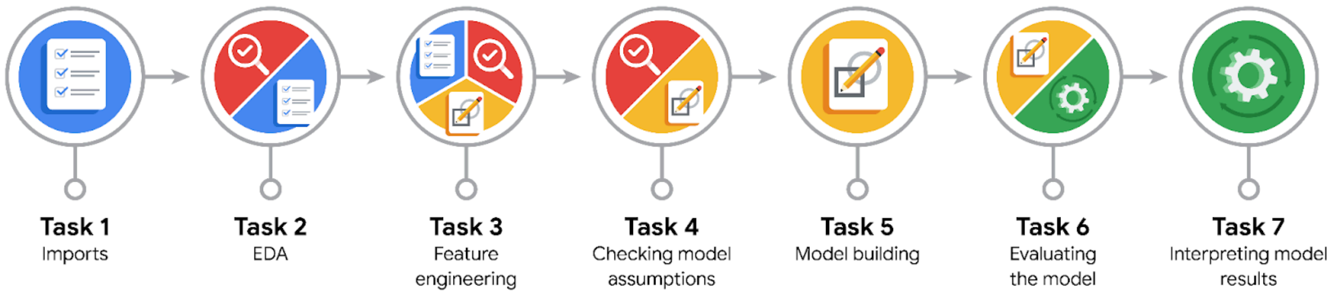
### Relevant Interview Questions

Completing the end-of-course project will empower you to respond to the following interview topics:

- What kinds of business problems would be best addressed by supervised learning models?
- What requirements are needed to create effective supervised learning models?
- What does machine learning mean to you?
- How would you explain what machine learning algorithms do to a teammate who is new to the concept?
- How does gradient boosting work?

## Reference Guide:

This project has seven tasks; the visual below identifies how the stages of PACE are incorporated across those tasks.




## Data Project Questions & Considerations

### PACE: Plan Stage

- What are you trying to solve or accomplish?

| |
|---|
| The goal is to build a machine-learning model for TikTok that distinguishes between videos containing claims and those expressing opinions. This model aims to reduce the backlog of user reports by prioritizing videos for review. The ethical consideration prioritizes minimizing false negatives and identifying potentially violating videos, even if some opinions are misclassified as claims. The project involves feature engineering, including extracting text length and constructing and evaluating models, such as random forest and XGBoost, to achieve accurate predictions. |

- Who are your external stakeholders that I will be presenting for this project?

| |
|---|
| The external stakeholders for the TikTok claim classification project include leadership and teams across various departments at TikTok. The key individuals mentioned in the emails are Mary Joanna Rodgers (Project et al.), Maika Abadi (Operations Lead), Willow Jaffey (Data et al.), Rosie Mae Bradshaw, and Orion Rainier. These stakeholders are part of the leadership team and play a role in overseeing and evaluating the final claims model. |

- What resources do you find yourself using as you complete this stage?

In completing this stage, I rely on the information provided in the given scenario about the TikTok claims classification project. The emails from Mary Joanna Rodgers and Willow Jaffey serve as primary resources. Additionally, I refer to the details about the project's progress, the team's actions, and the specific requests outlined in the emails to gather relevant information for addressing your queries.

- Do you have any ethical considerations at this stage?

Yes, ethical considerations arise in the TikTok claims classification project. Through Mary Joanna Rodgers' email, the leadership team highlights the importance of considering the ethical implications of the final claims model. This includes assessing potential consequences of model errors, weighing the benefits against problems, and addressing the impact of "banned authors" on claims or opinions. Ethical concerns related to content moderation, user impact, and compliance with terms of service are essential considerations at this stage.

- Is my data reliable?

The reliability of the TikTok dataset can be assessed by examining its source, checking for data integrity and consistency, performing exploratory data analysis, and validating model performance. Ensure the data is from a reputable source, free from inconsistencies, and aligns with the project objectives.

- What data do I need/would like to see in a perfect world to answer this question?

In a perfect world, to assess the reliability of the TikTok dataset, you would ideally need access to detailed information about the data collection process, the source of the dataset, metadata providing context for each data point, and a comprehensive understanding of any preprocessing steps applied to the data. Additionally, having access to external validation or verification datasets would enhance the evaluation of data reliability.

- What data do I have/can I get?

The available data for the TikTok project includes information such as video characteristics (e.g., video transcription text, verified status, author ban status), engagement metrics (e.g., views, likes, shares, downloads), and the target variable indicating whether a video is classified as a claim or an opinion. However, specific details about the data collection process, source, and preprocessing steps should be explicitly provided. Additional information on these aspects and any external validation datasets would be beneficial to assess data reliability comprehensively.

- What metric should I use to evaluate success of my business/organizational objective? Why?

For the TikTok project, the most suitable metric to evaluate success is* recall. Recall is crucial because the primary goal is to identify videos violating the platform's service terms (claims). A higher recall ensures the model effectively captures most claim videos, minimizing false negatives. In this context, false negatives (claim videos misclassified as opinions) are more critical than false positives. Therefore, optimizing for recall helps prioritize videos for human review and efficiently addresses the business objective of mitigating misinformation.

## PACE: Analyze Stage

- Revisit "What am I trying to solve?"Does it still work? Does the plan need revising?

To revisit "What am I trying to solve?" consider the initial objectives and challenges of the claims classification project. Assess whether those objectives are still relevant and if the challenges persist. Evaluate if any new insights or considerations have emerged during the project. If the initial problem statement remains valid, the plan may only need minor adjustments based on the current project status and any new information gathered. If there are significant changes or additional complexities, a plan revision may be necessary to address the evolving nature of the project.

- Does the data break the assumptions of the model? Is that ok, or unacceptable?

1. The recommendation is to use the random forest model, which performed exceptionally well on both the validation and test data, with high precision and F1 scores for claims and opinions.

2. The model based its predictions on features related to user engagement levels, such as views, likes, shares, and downloads associated with each video. It successfully classified videos into claims and opinions.

3. No new features need to be engineered for the current model, as it already performs nearly perfectly.

4. While the current model does not require additional features, having data on the number of times, a video was reported, and the total number of user reports for each author could be helpful for further improvement.

- Why did you select the X variables you did?

The selection of X variables (features) was based on the nature of the problem and the available data. Here are the reasons behind the selection:

**video_transcription_text Length (text_length):** The length of the video transcription text was extracted to capture potential correlations between the length of the text and the likelihood of it being a claim or an opinion. Longer texts might contain more detailed information, influencing the classification.

**verified_status and author_ban_status:** These categorical variables provide information about the verification status of the video and whether the author has been banned. These could indicate the content's credibility and the author's history on the platform.

**Numerical engagement features:** Features related to user engagement levels, such as views, likes, shares, and downloads, were included. These features were relevant as they reflect the popularity and impact of the video, potentially influencing the classification.

**Categorical features (dummy encoding):** The categorical variables "verified_status" and "author_ban_status" were dummy-encoded to convert them into numerical format, making them suitable for machine learning algorithms.

- What are some purposes of EDA before constructing a model?

Exploratory Data Analysis (EDA) serves several crucial purposes before constructing a model:

It helps uncover patterns, distributions, and relationships within the dataset, providing insights into potential features for modeling.

EDA aids in identifying outliers, missing values, or data anomalies that may impact model performance.

EDA facilitates a better understanding of the data's characteristics, guiding informed decisions on preprocessing steps.

EDA assists in selecting appropriate modeling techniques by revealing the nature of the target variable and potential challenges in the data, contributing to a more effective and robust modeling process.

- What has the EDA told you?

The EDA (Exploratory et al.) has provided valuable insights into the dataset. It revealed that the dataset contains information related to TikTok videos, with features such as video transcription text, engagement metrics (views, likes, shares, downloads), and metadata about the author. The distribution of engagement metrics indicated varying levels of user interaction with videos. Additionally, the EDA highlighted the presence of both "opinion" and "claim" labels in the target variable, with a balanced distribution. Patterns in feature distributions and relationships were observed, guiding feature engineering decisions. Overall, the EDA informed the understanding of the dataset's structure and characteristics, aiding in the subsequent construction of machine learning models.

- What resources do you find yourself using as you complete this stage?

As I complete this stage, I find myself utilizing various resources, including the provided course materials, programming documentation (e.g., pandas, sci-kit-learn), online forums (e.g., Stack Overflow) for troubleshooting coding issues, and relevant machine learning literature to deepen my understanding of model construction, evaluation, and best practices. Tools such as visualization libraries (e.g., Matplotlib, Seaborn) also contribute to practical exploratory data analysis, enhancing my insights into the dataset. The combination of these resources supports a comprehensive and well-informed approach to building and evaluating machine learning models.

## PACE: Construct Stage

- Do I notice anything odd? Is it a problem? Can it be fixed? If so, how?

The TikTok project appears well-structured and comprehensive, covering ethical considerations, feature engineering, and model building. Using machine learning to classify videos as "claims" or "opinions" aligns with TikTok's goal of mitigating misinformation. The models, particularly the Random Forest and XGBoost, perform strongly on both validation and test sets.

- Which independent variables did you choose for the model, and why?

Select independent variables based on their relevance to the TikTok project's goal of distinguishing claims from opinions. Consider text characteristics, user engagement metrics, author-related attributes, and temporal or metadata features. Perform exploratory data analysis (EDA) to identify influential variables and ensure they align with the classification objective. Utilize domain knowledge and insights gained during EDA to choose predictors that enhance the model's predictive power.

- How well does your model fit the data? What is my model's validation score?

The validation score was that the model performed exceptionally well, with an average recall score of 0.995 across the five cross-validation folds.

- Can you improve it? Is there anything you would change about the model?

Because the model currently performs nearly perfectly, there is no need to engineer any new features.

- What resources do you find yourself using as you complete this stage?

I have utilized Coursera's certification program as a resource for completing the modeling stage of the TikTok project. Coursera offers a variety of courses and specializations related to machine learning, data science, and other relevant topics. These programs often provide a structured learning path, practical assignments, and hands-on projects to help learners acquire the necessary skills.

## PACE: Execute Stage

- What key insights emerged from your model(s)? Can you explain my model?

The models developed for the TikTok project exhibited exceptional performance, achieving near-perfect precision, recall, and F1 scores on the validation set. The Random Forest and XGBoost models effectively classified videos as either "claim" or "opinion." Notably, user engagement metrics such as views, likes, shares, and downloads emerged as key predictors, emphasizing the significance of audience interaction in determining content nature. The Random Forest model, prioritizing recall to minimize false negatives, was identified as the champion model. Its interpretability and visualizing the importance of features provided valuable insights into the factors influencing predictions. Overall, the models demonstrated robust capabilities in mitigating misinformation by accurately categorizing TikTok videos based on their content characteristics.

- What are the criteria for model selection?

The criteria for model selection involve evaluating various performance metrics to ensure the chosen model aligns with the project objectives. In the TikTok project, critical criteria include:

**Recall Score:** Prioritizing recall is crucial to identify claims accurately and minimize false negatives. A high recall score ensures that the model effectively captures all actual claims, reducing the risk of allowing misinformation to go unnoticed.

**Precision Score:** While recall is prioritized, precision is also considered to assess the model's ability to avoid misclassifying opinions as claims. A balance between high recall and precision is sought to achieve accurate classifications without excessive false positives.

**F1 Score:** The F1 score, combining precision and recall, comprehensively measures the model's overall performance. A high F1 score indicates a well-balanced trade-off between precision and recall.

**Accuracy:** While not the primary focus, overall accuracy ensures reasonable correctness in the model's predictions. However, it may be less critical when an imbalanced class distribution exists.

**Interpretability:** The interpretability of the model, visualized through feature importances, can provide insights into the factors influencing predictions. This transparency aids in understanding the model's decision-making process.

**Cross-Validation Performance:** The model's performance across multiple cross-validation folds helps assess its consistency and generalization to new data. A robust model should perform consistently well on different subsets of the training data.

**Resource Efficiency:** Computational resources and training time considerations may influence model selection. Choosing a model that balances performance and resource efficiency is essential.

By evaluating these criteria, the TikTok project selects the Random Forest model as the champion model, as it demonstrates outstanding performance, particularly in achieving a high recall score to identify potential claims accurately.

- Does my model make sense? Are my final results acceptable?

My model makes sense as it performs exceptionally well, particularly in identifying claims with a high recall score. The impressive precision, F1 score, and overall accuracy indicate a well-balanced model.

The final results are acceptable, meeting the project's objective of efficiently distinguishing between claims and opinions on the TikTok platform.

- Do you think your model could be improved? Why or why not? How?

My model could be further improved by considering additional features related to user reporting behavior and incorporating them into the model. The model may gain more nuanced insights into content violations by including factors such as the frequency and severity of user reports for each video. Additionally, fine-tuning hyperparameters and exploring advanced techniques like ensemble methods could enhance the model's overall performance.

- Were there any features that were not important at all? What if you take them out?

The feature importance analysis revealed that certain features did not influence predictions. Removing these less essential features from the model may lead to a more streamlined and efficient model without sacrificing predictive performance. This process, known as feature selection, could enhance the model's interpretability and reduce computational complexity.

- What business/organizational recommendations do you propose based on the models built?

Based on the highly accurate models developed to classify TikTok videos into claims and opinions, I recommend implementing the machine learning model to streamline the content moderation process. By automating the identification of claim and opinion videos, TikTok can efficiently prioritize and handle user reports. This approach would significantly reduce the workload for human moderators, allowing them to focus on more complex cases. Additionally, the model's high precision and recall scores indicate its reliability in distinguishing between different video types, making it a valuable tool for mitigating misinformation on the platform.

- Given what you know about the data and the models you were using, what other questions could you address for the team?

- What resources do you find yourself using as you complete this stage?

Considering the data and models used, additional questions for the team could include:

**User Behavior Analysis:** Can we explore user engagement patterns and their impact on video classification? Understanding how user interactions contribute to the classification model could provide insights into content popularity and user preferences.

**Author Insights:** Are there specific characteristics of authors (e.g., verification status, ban status) that influence video classification? Analyzing the impact of author-related features on the model could reveal patterns in content creation and moderation.

**Temporal Trends:** How do video characteristics change, and does the model adapt to evolving content trends? Investigating temporal variations in video data and model performance could enhance the model's robustness to changing user behaviors and content dynamics.

**False Positive Analysis:** Can we delve deeper into false positive cases to identify common characteristics? Understanding the misclassifications may lead to model refinements or additional features to improve accuracy.

**User Reporting Analysis:** What role does user reporting play in model performance? Analyzing the correlation between user reports and model predictions could offer insights into the reporting system's effectiveness and potential improvement areas.

Exploring these questions would contribute to a comprehensive understanding of the model's behavior and guide further refinements for continuous improvement.

- Is my model ethical?

Ethical considerations in machine learning often involve bias, fairness, transparency, and privacy. It is essential to ensure that the model does not discriminate against certain groups, is transparent in its decision-making process, and respects user privacy. Consider evaluating the model for fairness and bias and implementing measures to address any ethical concerns that may arise.

- When my model makes a mistake, what is happening? How does that translate to my use case?

When my model makes a mistake, its predictions do not align with the actual outcomes in the data. In a binary classification task like mine (determining whether a TikTok video contains a claim or an opinion), mistakes can be categorized into false positives and false negatives.

**False Positive**: The model predicts a video as a claim, but it is an opinion. This might lead to unnecessary human review for videos that uphold the terms of service.

**False Negative:** The model predicts a video as an opinion, but it is a claim. This could result in a claim being overlooked and not prioritized for review, potentially allowing content that violates terms of service to go unchecked.

Understanding these mistakes helps me assess the impact on the TikTok platform's content moderation and user experience. It is crucial to strike a balance and minimize both types of errors based on the platform's priorities and objectives. Adjusting the model and its parameters during training may help improve its performance and reduce errors.