

Case Study Rubric: Does Emotional Sentiment Assist Spam Email Classification?

Submission Format: About 2 pages in PDF format

Why am I doing this?

This case study is your opportunity to get familiar with vital data science skills, including exploratory data analysis, feature engineering, machine learning, statistical testing, and data communication, and apply it to a real-world problem. You will investigate whether the emotional sentiment of an email helps predict whether it is a phishing attempt.

You will build a machine learning model, evaluate its effectiveness, and communicate your findings through visualizations and a professional report. This project prepares you to clearly present technical work to a broader audience and think critically about modelling decisions.

Learning Objective: Prepare findings for presentation to your peers.

What am I going to do?

You will complete a structured case study designed to familiarize you with sentiment analysis and text-based machine learning.

You will:

- Work through provided exploratory data analysis (EDA) and modeling notebooks.
- Carefully review the steps taken in loading data, engineering features (TF-IDF and sentiment), training a model, and testing significance.
- Create meaningful visualizations based on the results.
- Write a short, professional report summarizing your understanding of the case study.

Tips for Success

- Focus on understanding the workflow. Think critically about the results.
- Prioritize clarity over complexity, and focus on writing for a general audience.
- Use visualizations thoughtfully. Choose graphs that help tell the story of your analysis.

Spec Category	Spec Details
Formatting	<ul style="list-style-type: none">• Submit a polished PDF report.• ~2 pages, TNR 12-point font.• Your report should include the following sections:<ul style="list-style-type: none">○ Project Overview○ Data and Modeling○ Results○ Conclusion

Project Overview	<ul style="list-style-type: none"> • <u>Goal:</u> this short paragraph should help the reader understand the scope of the case study. • Summarize relevant ideas and the topic at hand. • Indicate which contents of the case study were relevant for your project.
Data and Modeling	<p>Data Paragraph</p> <ul style="list-style-type: none"> • <u>Goal:</u> describe the data from the case study. • Identify the relevant features and outline the EDA. Add any EDA if you would like. • Discuss the class balance between phishing and safe emails. <p>Modeling Paragraph</p> <ul style="list-style-type: none"> • <u>Goal:</u> describe the modeling approach from the case study • Describe the feature engineering and modeling choices. How did they contribute to the results?
Results	<ul style="list-style-type: none"> • <u>Goal:</u> briefly describe the results from the case study • How did the model perform? What features were powerful predictors?
Conclusion	<ul style="list-style-type: none"> • <u>Goal:</u> describe what you learned from the case study • Reflect on at least one new thing you learned. This could be an organizational approach, visual technique, or anything else. • Is there any part of this case study that you could incorporate into your work? • Is there any element that might not be a good idea to incorporate into your work?