

Stratifyd Data Science Assignment (Sentiment Prediction V)

Main Task

1. Clean and rearrange the dataset in the attachment.
2. Use Semi-AAE to design a classification model, which is able to predict whether the customer's review is positive or negative.

The format of the submission should be a script that can receive arguments from the terminal and generate desired outputs.

Data

- pos.txt: positive reviews from an online shopping website
- neg.txt: negative reviews from the same online shopping website

Specific Requirement

Output: Whether a customer's review is positive or negative;

Input: Any information in the dataset.

Code: You can refer to any sample code if needed, but remember to give the reference when you did. Note that you still need to know **exactly** how the code works. The referenced code should be **less** than **50 %** of the whole script.

Model: You should know **every detail** of the models you used, including the principles, how to fine tune the hyperparameters, etc.

Extra Credit

1. Think about if there is any trick or technique to improve the predicted results.
2. Implement the semi-supervised classification model proposed in the attached paper, compare the prediction accuracy (f1 score) with the previous model.

Submission format

A project folder that can handle arguments and generate desired outputs. The script should also include comments for each key block. A short **PDF** report that contains all program outputs regarding the model.

Note

The purpose of this challenge is to evaluate your background. If you **couldn't fully** finish the challenge, please send us:

1. The code you have finished so far. (You have to clean it up and provide comments if needed)
2. A short report of this challenge, e.g., your progress so far and the issues you haven't solved yet, etc.

Based on the effort you have shown in the challenge, we will let you know if you are qualified to pass this challenge and setup a 2nd round interview with you to discuss details of the assignment and your experiences on CV.