



Take Home Assignment: Data Scientist

Thank you for your interest in StyleSeat and we appreciate you taking time to work through this exercise.

Please solve the following problem using the tools and libraries of your choice. You are not expected to spend more than 3 hours solving this take home assignment. We are not expecting a perfect solution, but want to understand your approach and how you've gone through the process to solve the exercise as well as the deliverable. It is quite normal for candidates to reach out with specific clarifying questions via the recruiter.

Expected deliverables are listed at the end of this document.
(use of model).

Problem : Text Classification

Attached with this problem set is a file called *training_data.csv* which contains a matrix with two columns: *service_name* and *service_category*. Each row in this matrix is a service that can be found on StyleSeat.com. The column *service_name* contains the name of the service, while the column *service_category* lists the categories (separated by a comma) the respective service falls under.

Train a classifier that predicts the category of a service given its name. In doing so answer the following questions:

- A. Is this a multi-label classification problem? Briefly explain the implications.
- B. Show and explain any data processing/feature engineering performed.
- C. What ML model did you use and why?
- D. Explain any measures you took against overfitting.
- E. How did you assess the efficacy of your model?
- F. Run the model against the data in *testing_data.csv* and provide the results in *yourname_submission_problem_1.csv*
- G. What further enhancements to the model would you consider if you were given more time?

Deliverables :

- Brief write-up answering A-G above. We may use this as a verbal discussion point.
- Supporting code (.py .ipynb or any other).
- Your trained model (.pickle, .h5 or any other).
- File `yourname_submission_problem_1.csv`