

Senior DS / ML Engineer - Home Assignment

Introduction

Our legal department has completed defining domain names according to specific legal laws. They have labeled manually a dataset of legal cases and assigned a match domain to each case. In the future, we want to utilize this labeled dataset and assign a domain to every new legal case we acquire.

You are expected to deliver a classifier that, given a legal case document, predicts the correct domain.

Expected DS Deliverables:

1. Exploratory Data Analysis (EDA) of the given data including relevant insights and visualizations that helped you analyze the dataset and choose the best fit solution.

Make sure to include analysis on the following:

- Inconsistencies or errors in the data
- Descriptive statistics

2. Attached Jupyter notebook containing the algorithm you chose for the challenge, including all relevant aspects of the process.

Follow and use the existing methods in the notebook and add more as needed.

3. Metrics for your solution including an explanation of which metrics are more important and how you optimized them.

4. Summary document with the following:

- a. What did you achieve?
- b. What performance metrics did you get?
- c. If you had more time, what would you do differently?