ML Engineer Take Home Task

Classification of Google Patents

This assignment is designed to give you a hands-on opportunity to work on a problem related to work we do here at Govini. If you enjoy this kind of challenge, you will love working with our team.

Instructions

Use a subset of the publicly available Google patents dataset to create a binary classifier that classifies each row as 0 or 1.

Data

The parquet file can be downloaded here:
https://s3.amazonaws.com/BUCKET_FOR_FILE_TRANSFER/ml_dataset.parquet

Packages

We use the following libraries, but you can choose whatever libraries you like to complete this task:
- Huggingface Transformer
- Huggingface Accelerate
- Huggingface Evaluate
- Polars
- Pandas
- loguru
- Hydra