

Take-Home Machine Learning Engineer

This take-home aims to assess your skills as well as providing you with insights into the kinds of data projects at Levi’s. The question attempts to cover many of the necessary skills of an entry level ML engineer including software engineering and insight generation.

To help you plan, the questions/tasks can be answered/completed within 4-6 hours.

Please contact me if any questions [psalvadorlopez@levi.com](mailto:psalvadorlopez@levi.com).

Background:

**Goal**: Automate an ETL pipeline using software/data engineering best practices and evaluate your insights generation.

**About the Data**: We are asking you to use the provided banking dataset to perform this exercise.

**Problem/Tasks:**

1. *Data/Software Engineering:* 
   1. Automate an ETL (Extract, Transform, Load) pipeline, using Python. Please, include the following steps in your pipeline as a requirement, however you are welcome to add any other method you find necessary to prepare the data to be consumed by an ML algo:
      1. Drop all the high correlated columns with the Attrition Flag column (if any)
      2. drop null or incorrect values (if any)
      3. apply one-hot or Ordinal encoding in categorical features
   2. Structure your code as reusable classes with appropriate unit tests
2. *Data Science:*
   1. Identify outliers in the data, justify your choice of statistical outlier detection.technique
3. *CI/CD*:
   1. Create a Github Repository to store and track all your code for this exercise
   2. Create a new branch in the GitHub repository called “dev”
   3. Initiate a pull request from “dev” branch to “prod” once your work is completed done.

Looking forward to hearing back from you.