CWT Python Developer - Case Study

**Context**

For this case study, you will use this dataset available on Kaggle. Please download it from this link: <https://www.kaggle.com/datasets/nicapotato/womens-ecommerce-clothing-reviews>

This dataset contains reviews written by customers on clothing. Please download the Womens Clothing E-Commerce. Please note that this data is released by academic researchers. It is not used commercially at CWT.

You will also find a Python Notebook on the email attachment called CWT Python Developer Case study. This notebook contains a basic codebase for pre-processing the data from the Womens Clothing CSV, train a basic Logistic Regression model and perform GridSearch to identify the best hyperparameter that yields the highest score. You are only required to work on the following tasks, so there is no need to spend time re-training the model.

**Tasks**

1. Refactor the code that is listed under **Task 1** in the notebook. You can find instructions there as well. Please refactor the code and create a base class and other classes that will inherit and implement the functions mentioned in Task 1.
2. We would like to increase the speed of the function called grid\_search in **Task 2**. Please read the instructions under **Task 2** in the notebook.
3. Using the data provided in the CSV file, please create a dummy DB with SQLite and a RESTful API with CRUD operations on the primary key or document ID. (You can use any database of your choice (SQL or NoSQL) but please provide documentation on how to reproduce this in our local environments). Your schema should reflect the same schema as the CSV file.
4. On the attachment you can find a file called model.pkl. Please create an endpoint called inference to serve the predictions of the model.
5. **[OPTIONAL**] Pleasebuild your Python project as a package.

**Constraints**

* Please provide the case study back as a GitHub repository
* Make sure you write some unit tests
* Make sure your code contains documentation and we can reproduce it and run it on our environments

**Please return your solution within 1 week.**



© 2022 CWT