Graded Assignment

Week 6

User Story:
You are working within a clothing firm as a data scientist. You are responsible for training and deploying the models that you create. The firm has to be able to predict trends as well as they can because they need to produce the said clothes beforehand and supply chain issues need to be resolved. Due to the critical nature of the task, it is imperative that you get the demand forecasting right.

Problem Statement:
But as is the nature of the business, the fashion industry moves fast and no one trend remains for long, this renders your best trained forecasting models useless after some time. To resolve this, the team needs to constantly update and retrain their models all the time. The data you are working with has the following columns :-

1. **URL** - URL of the Product
2. **Product_id** - ID of the Product
3. **BrandName** - Product Brand Name
4. **Category** - Category of product
5. **Individual_category** - Sub category of Each product
6. **category_by_Gender** - Category based on gender
7. **Description** - Description about the project
8. **DiscountPrice** (in Rs) - Price after Discount
9. **OriginalPrice** (in Rs) - Original Price of the Product
10. **DiscountOffer** - Offer on the product
11. **SizeOption** - Different Size Options
12. **Ratings** - Customer Ratings
13. **Reviews** - Total Customer Reviews

User Stories:

- You have to create a model that can predict user ratings.
- Customer behavior changes over time so you need to showcase the ability to retrain the model and showcase the effectiveness of this approach
- You want to showcase utilization of multiple models and how they are affected by multiple trainings
- Do they perform better, worse or the same. Document and analyze the same extensively to showcase to the management.

Instructions :-

- Please write your code in jupyter notebook with all the steps

- Use markdown extensively and showcase if model retraining had any impact, do not limit yourself to using one algorithm, use as many as you can to showcase if model retraining has any impact
- You are expected to submit your
    - Jupyter notebook as HTML
    - A word document that will be your report to the management which will have a
        - Summary of the problem statement
        - Your approach towards a solution,
        - What all you tried
        - How well it worked
        - Your analysis and recommendations for the solution

Submission Guidelines:

Notebook Submission:

- Submit a Jupyter notebook containing the code for training models, retraining models.
- Use markdown extensively
- Ensure clear documentation on how to reproduce the results.

Word Document Submission:

- Provide a Word document containing:
- Charts from the notebook, including visualizations of model comparisons and hyperparameters.
- Detailed analysis, findings, and insights derived from the trained models.
- Screenshots backing up analysis and claims made in the report.
- A section highlighting the input and recommendations for management.
- Consider incorporating user stories to convey the user perspective in the analysis.

Checklist for Word Document:

- One should be able to see how many models were tried.
- One should be able to see which models were tried.
- One should be able to see all the hyperparameters that were tried.
- The scores of the models should be visible.
- The scores of the models should be compared in charts for easier comparisons.
- A report at the end giving input as per the analysis, considering the user stories.