codemonk.

Machine Learning Internship Assignment

Philosophy

The goal of this challenge is to showcase your coding ability, analytical thinking, and system-building capability and last but not least demonstration of strong Data Science skills.

Values

- We use Python programming language extensively
- We love getting rich insights from data and always do detailed EDA
- We prefer Pytorch but you can still use Keras or Tensorflow
- We highly value code cleanliness, re-usability, structure and proper documentation.

Time Limit

Complete this challenge within 2 days of receiving it.

The Challenge

Task 1:

Download the following dataset

Dataset Link: https://data.world/thatzprem/agriculture-india

Using the dataset carry out the following task:-

1. Get interesting insights from the dataset (detailed EDA in Jupyter Notebook)

codemonk.

Machine Learning Internship Assignment

Task 2:

Download the following dataset

Dataset link: https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-dataset

Using the dataset carry out the following tasks:-

1. Detailed EDA
2. Given an image, input build a deep learning based model that can predict the following outputs:-
    a. Color of product
    b. Type of product (T-shirt, shoes, etc.)
    c. Preferable season to use this product
    d. Gender (Men, Women, Unisex)
3. Once you build the model, take a few sample screenshots of fashion products from the Amazon website and use your model to predict the result.

Bonus points

1. If you can create an API script for serving the model
2. Create Streamlit or any other GUI-based application for demonstration

codemonk.

Machine Learning Internship Assignment

Note:

1. For task 2 you can directly create a notebook in Kaggle. Train the model using GPU compute available. Once you have trained the model, download the Jupyter Notebook. And after training depending on model size, use some storage like a google drive.
2. You can make certain valid assumptions to do the task efficiently but make sure you write it in a notebook explicitly.
3. Jupyter notebook, trained model link and miscellaneous required files for running code should be made available in Gitlab or Github

Submission:

When you are all done:

1. Make a public Github or Gitlab repository and upload all task files in it.
2. Make sure we can run the code with ease.
3. Shoot us a quick email letting us know once you're done.

All the best!