Take Home Assignment

Introduction

Thank you for your interest in the Machine Learning Engineer role within the IT – AI/ML Team at SHI International Corp. As part of the interview process, we would like you to complete a take home assignment. After you have completed the assignment, we will review the deliverables and schedule a follow-up technical interview, if needed.

Scenario

You are a Machine Learning Engineer. You have been asked to create a machine learning model to categorize products. The product owner for this project has obtained a labeled dataset for you to develop a proof-of-concept. What will you do?

Instructions

- Write a design specification document that summarizes the problem, assumptions, solution, alternative solutions, and open questions.
    - Design should address data preprocessing, ML model selection, training, hyperparameter tuning, testing, metrics, deployment, monitoring, maintenance.
    - Include diagrams as needed.
- Create a repository on GitHub and write the code for your solution using standard coding practices.
    - Code should be written in Python.
    - Platforms and/or packages that you decide are fit for purpose may be used.
    - Coding assistants may not be used.
    - Include instructions for environment setup and running the code.
- Deliverables
    - Design specification document (PDF format)
    - Link to GitHub repo
    - Logs from successfully run code (PDF format)
    - Summary of ML model performance metrics (PDF format)
    - Predictions on un-labeled dataset (CSV format with columns "Index" (0, 1, 2, 3, etc.), "Manufacturer", "SKU", "PredictedCategory", "PredictionScoreOrProbability")

Datasets

- products_labeled.json.gz: Product records with "Category" label.
- products_unlabeled.json.gz: Product records without "Category" label.

Sample from products_labeled.json.gz

```
[
  {
    "Category": "Sports & Outdoors",
    "Price": null,
    "Title": "Ultra Pro Gaming Generic Deck BoxDeck Box, Multi, One Size",
    "Features": [
      "Premium high-gloss design with a metallic finish",
      "Holds 100+ standard-size cards with or without deck protectors",
      "Spacious accessory compartment at the bottom, perfect for storing dice & other accessories.",
      "Country of origin: United States"
    ],
    "Description": [
      "Our Ultra Pro Hi-Gloss Satin Towers are designed to hold more than 100 Standard Size Trading Cards in Ultra PRO Deck
Protectors Sleeves. Designed with smooth satin and a metallic finish; this premium deck box has an accessory compartment in the
bottom which is great for holding life counters, dice, and other gaming necessities.)"
    ],
    "Details": {
      "Item Package Dimensions L x W x H": "5.12 x 3.39 x 3.19 inches",
      "Package Weight": "0.23 Kilograms",
      "Item Dimensions  LxWxH": "3.15 x 3.15 x 5.12 inches",
      "Item Weight": "0.5 Pounds",
      "Brand Name": "Ultra Pro",
      "Model Name": "Deck Box",
      "Color": "Multi",
      "Material": "Other",
      "Suggested Users": "Unisex-adult",
      "Number of Items": "1",
      "Manufacturer": "Caseys-Distributing",
      "Part Number": "85414",
      "Style": "Deck Box",
      "Included Components": "1 Item",
      "Size": "One Size",
      "Best Sellers Rank": {
        "Toys & Games": 745886,
        "Collectible Trading Card Cases": 1440
      },
      "Date First Available": "October 9, 2017"
    },
    "SKU": "85414",
    "Manufacturer": "Caseys-Distributing"
  },
  {
    "Category": "Fashion",
    "Price": null,
    "Title": "Iron Fist Women's Zombie Stomper Fugg Boot",
    "Features": [
      "100% Synthetic",
      "Manmade sole",
      "Vegan"
    ],
    "Description": [],
    "Details": {
      "Package Dimensions": "12 x 8 x 4 inches; 2 Pounds",
      "Item model number": "Zombie Stomper Fugg",
      "Department": "Womens",
      "Date First Available": "November 16, 2011",
      "Manufacturer": "Iron Fist"
    },
    "SKU": "Zombie Stomper Fugg",
    "Manufacturer": "Iron Fist"
  }
]
```

Citation

This is a very small, slightly transformed, and slightly fabricated sample of data from https://amazon-reviews-2023.github.io/ .

```
@article{hou2024bridging,
  title={Bridging Language and Items for Retrieval and Recommendation},
  author={Hou, Yupeng and Li, Jiacheng and He, Zhankui and Yan, An and Chen, Xiusi and McAuley, Julian},
  journal={arXiv preprint arXiv:2403.03952},
  year={2024}
}
```