**Clothing Similarity Search**

**Overview:**

The goal of this project is to create a machine learning model capable of receiving text describing a clothing item and returning a ranked list of links to similar items from different websites. Your solution must be a function deployed on Google Cloud that accepts a text string and returns JSON responses with ranked suggestions.

**Steps:**

1. Collect and preprocess data.

-Use web scraping tools to gather a dataset of clothing item descriptions and their corresponding URLs from multiple e-commerce websites. Ensure that the dataset is balanced and diverse, including various types of clothing from different brands and retailers.

-Preprocess the text data by cleaning it (remove special characters, lowercasing, etc.), and possibly by applying some form of text normalization (like stemming or lemmatization).

2. Measure similarity.

-Develop a method for extracting useful features from the text descriptions. This could involve using techniques like TF-IDF, word embeddings (Word2Vec, GloVe), or more advanced techniques like transformer-based models (BERT, GPT).

-Implement a method to compute the similarity between the input text and the texts in your database. This could be done using methods like cosine similarity or Jaccard similarity.

3. Return ranked results.

-Design a function that accepts a text string, extracts its features, computes similarities with the items in the database, ranks them based on similarity, and returns the URLs of the top-N most similar items.

4. Deploy the function.

-Containerize your application using Docker, then deploy your function on Google Cloud Functions or Google Cloud Run. Your function should accept a JSON payload with the input text and return a JSON response with the ranked list of similar item URLs.

**Rubric:**

1. Completeness. Did the engineer implement all components of the project and carefully follow instructions?

2. Performance. How well does the project achieve its intended goal? Is the implementation efficient?

3. Documentation. How easy is the code to understand? Is code well-commented? Is there a ReadMe file with an overview of the project and instructions for deployment?