Food Word Embedding and Clustering Task

Overview: The goal of this exercise is to cluster the provided textual data based on the semantic meaning of the words in the corpus provided.

Instructions: The task must be completed in Python. You can use the packages and algorithms you prefer. Report all results in Jupyter Notebook, uploaded to your GitHub page. All steps must be visible. Make it clear and simple. Please include comments, titles, and explanations. The task should take approximately 2h to finish. The deadline for this task is **Saturday June 25, 11:59pm PST**

1. **Download the *MenuItem* dataset provided at >>> this link <<<.** This dataset consists of **4,524** menu items from restaurant brands. Your goal is to *cluster together restaurant brands* based on their menu items being used in semantically similar contexts. We define the context of a restaurant brand to be all the menu items it sells.

2. **Perform any data preprocessing, if necessary, to improve your final results.**

3. **Train an embedding model:** Use the preprocessed dataset with a state-of-the-art (SOTA) embedding method of your choice (e.g., Word2Vec or BERT sentence embedding) to produce a vector representation of the data.

4. **Cluster semantically similar ingredients / restaurant brands:** Use a SOTA clustering method (e.g., K-means, DBSCAN, hierarchical) to cluster together semantically similar ingredients (if using the ***Recipe*** dataset) OR restaurant brands (if using the ***MenuItem*** dataset) based on their being used in semantically similar contexts. You may need to use dimensionality reduction on the vectors obtained from the embedding model before applying the clustering algorithm.

5. **Evaluating the clusters**: Use quantitative metrics to evaluate the quality of the clusters. Visualize the clusters in 1-2 interactive figures, including descriptive captions. Feel free to describe the clusters (e.g., sense-check them) for meaning.

6. **Reporting the results**: Make a Jupyter Notebook or RMarkdown document explaining all the steps you perform. Upload the results to your GitHub page. Make sure the repository is public and submit the link to the repository here: **LINK TO SUBMIT REPOSITORY**