**Motivation**

Food waste is a major issue in the United States. An American family of four discarded on average $1600 value of produce annually.1 Our project, Forage, is a machine learning algorithm that considers what you have in the fridge or pantry, to generate a unique recipe that utilizes those available ingredients.

**Data**

We filtered and extracted title, categories, ingredients and instructions of each recipe from the largest, free recipe database Meal-Master and picked out food nouns to concisely represent ingredients and simply tokenized instructions in sentences to preserve the specific “recipe language”.

**Features**

Word2Vec: Vector representation of words allowed to map semantically similar words close by each other.

We used Skip-Gram model that predicts source context-words from target words, which makes more sense on larger datasets and recipe evaluation.

**Results**

We first classified our dataset using k-means clustering, then assigned our generated recipes to the closest centroid. For initial testing, we used metric evaluations (Euclidean and cosine distance and Language Tool score), and we plan on using human evaluations to accurately determine the success of our model.

**Evaluation**

Out of the necessity for our model to understand the complex nature of our data (i.e. ingredients, steps, recipe titles, and categories) we designed and implemented a new approach to facilitate genuine and dynamic interaction with a trained model, which we call Model Interaction Language, or MIL. MIL is similar to the structure of JSON.

**Discussion**

Overall, our generated recipes made semantical sense and had an expected structure, thanks to MIL. However, the correlation of ingredients in the “ingredients” and “instruction” sections still needs improvement, as this relation was not characterized in the model cost function. Although LSTM model was computationally expensive to train, we found by controlling the number of most frequent word as vocabulary, the model could run faster and avoid overfitting. While standard evaluation is difficult to achieve in this context, we were able to use k-means and Language Tool to test our results, as they do not require a reference.

**Future Work**

We would like to conduct human evaluations on a larger scale to more accurately gauge the success of our model. We are continuing hyperparameter sweeps to optimize our model, including training with 512 hidden units in each layer. We will also explore PCA analysis to aid in model evaluation.

**References**
