ML Challenge Report

Siqi Liu Rafay Usman Ryan Shiels Albert Li

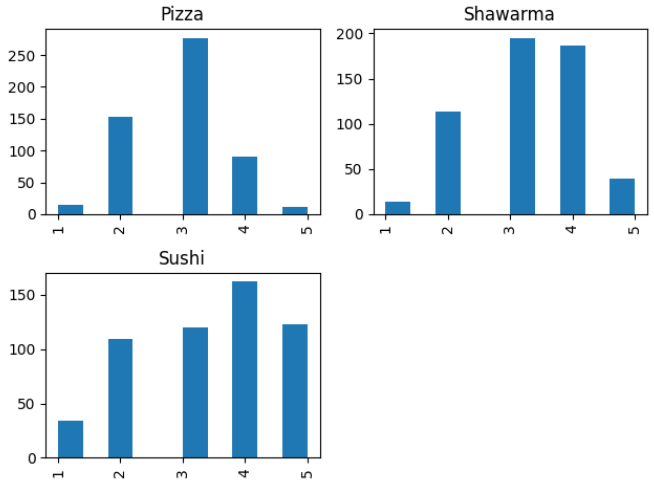

Fig. 1: Histogram for Q1

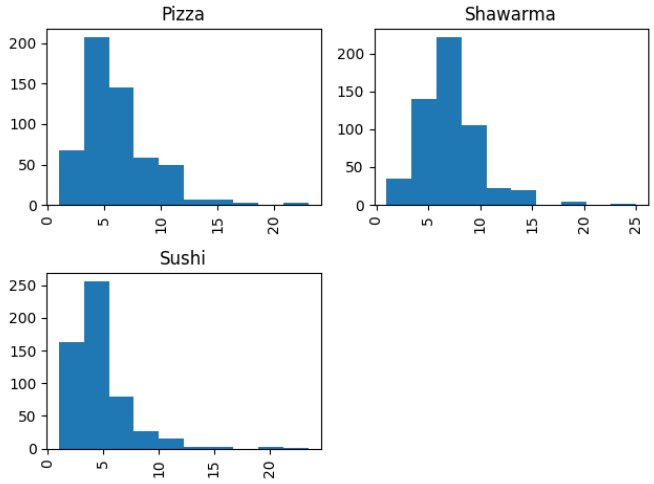

Fig. 2: Histogram for Q2

I. Data

We used a variety of plots to explore the data depending on the data type.

For numerical features such as Q1: How complex is it to make the food Q4: How much would you pay for the food, we used box plots and histograms to see the distribution of data for each food class. Using boxplots allowed us to identify the median and outliers of the responses, allowing us to see whether some food correlates with the question more than others.

We see an interesting correlation between the number of ingredients and the food. For pizza, we have a normal distribution where the majority of the inputs labeled it as having a difficulty of 3. For sushi on the other hand, the distribution is more uniformed, with a peak of 4 but a wider spread.

For "how many ingredients would you expect this food item to contain", we see the following plots showing a right skewed distribution for all food items. For "How much would you expect to pay for one serving of this item", we see the median for pizza is lower than sushi, and we see sushi with more extreme outliers than the other food classes.

We used bar charts for categorical data and only explored the questions with interesting data here. Some questions allowed open-ended answers, such as what movie associates with the food class. To make analysis simpler, we only plotted the top 5 most popular responses. The following figure shows the results for Q5: What movie do you think of when thinking of this food item?. For both Pizza and Sushi, we saw "none" as the most popular input, and interestingly we saw "Avengers" to be by far the most popular response for Shawarma, suggesting a correlation between Avengers and Shawarma.

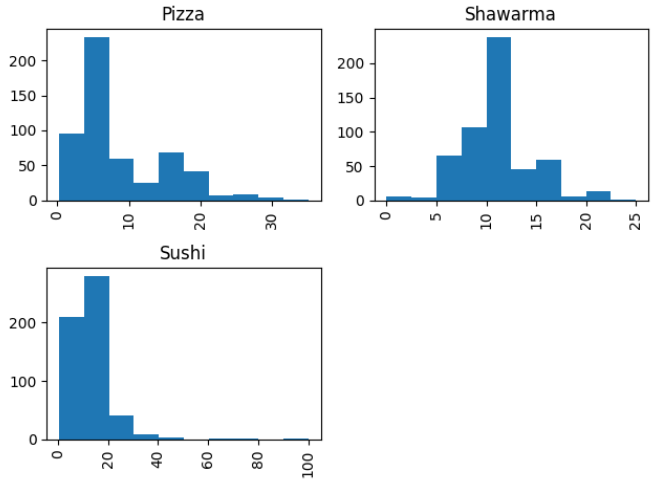

Fig. 3: Histogram for Q4

For Q3: "In what setting would you expect this food to be served?", we see a trend that Pizza is more appropriate for most situations, and shawarma and sushi are more specific in when they are expected to be served.

Although there are other features such as Q7: "when you think about this food item, who does this remind you of?", they show less differences for different food classes. For example, the most popular answer for Q7 was "Friends", making it less indicative of the food class.

We split the dataset into 3 sets: 60% training, 20% validation and 20% test. This allowed us sufficient data to train the model as well as data for testing that the models generalizes to unseen data

II. Model

A description of the model(s) that you are evaluating/exploring. We are expecting a thorough exploration of at least 3+ family of models1, even if you don't ultimately use those models. We are looking for: – How you are applying these models. You don't need to reiterate what the models are and how they work. Instead, we're looking for a description of the choices you are making to apply the model to this task: e.g., what features from the "Data" section are you using for each model? What adjustments (if any) did you need to make?

III. Model Choice and Hyperparameters

How you are determining what model(s) to use in your final pred.py, as well as an exploration of hyperparameters. We are looking for: – A convincing explanation of how you ensured that the evaluation metrics for various models are comparable (i.e., are you using a consistent test set?) – A clear description of the evaluation metric(s) used to evaluate your model, as well as justification for its use. – A description of the hyperparameters that you are tuning, hyperparameter value combinations that you have tried, and the value of the evaluation metric(s) for those hyperparameters. We are not looking for an exhaustive search of all possible hyperparameter combinations, but there should be enough evidence to demonstrate that your hyperparameter choices are reasonable. (Note that just saying "X was the best hyperparameter value based on metric Y out of the hyperparameter values we tried" is not enough. Please present the values of the evaluation metrics for each hyperparameter value to show that your choice was the best one.) – A clear description of what your final model choice looks like in the submitted pred.py file. – The descriptions should be consistent with the .py and/or .ipynb files that you used while developing your model.

IV. Prediction

How well would you expect your model to perform on the test set? We are looking for: – A point estimate for your performance, not a range. – A reasoned explanation of your expected model performance, with empirical evidence supporting your explanation. You are not graded on the closeness of your estimate to the final test accuracy, but you are graded on your reasoning.

V. Workload Distribution

A description of what each person in the group contributed to the project. A 1-2 sentence description of each person's role is sufficient. Each student's description must be written by that student in order for them to receive credit for the project.