

HOMEWORK-2: TEXT CLASSIFIER MODEL FOR RECIPES

Group 2: EMBEDDING EAGLES

Objective:

To determine whether a Reddit post or comment from r/Cooking is relevant to recipe writers, focus on recipe content, cooking techniques, ingredients, and culinary discussions.

Annotation Guidelines:

The first guideline for annotation is to define all entities. Defining is the most vital and basic step, as it helps and guides people with no prior knowledge to understand and perform annotations by giving them a basic understanding of what to do.

There are two categories for this project:

- **Relevant:** Content related to recipes that are relevant to recipe makers comes under Relevant. In our opinion, recipes contain ingredients, equipment, and the process of making a dish.
E.g., "For a classic Bolognese, I recommend using a mix of beef and pork for depth of flavor and adding a splash of milk towards the end for creaminess."
- **Irrelevant:** Content related to recipes irrelevant to recipe makers comes under Irrelevant. Anything unrelated to ingredients, equipment, or the procedure of cooking falls under the category of irrelevant.
E.g., "Pizza Hut is less expensive than Dominos."

Criteria for Relevance:

- Text, including recipes and procedures to cook a particular dish.
- Tips to follow for making certain recipes effortlessly and effectively.
- Description of equipment and ingredients regarding review of equipment, their usage to cook easily, and flavors to make delicious recipes.

Criteria for Irrelevance:

- Experiences or comments that do not provide insights into cooking techniques, ingredients, or recipes.
- Partial statements without including content about food and personal preferences.
- Conversations are unrelated to cooking or recipes.

Annotation Process: Manual annotation using Prodigy

- First, we read the description of comments entirely before deciding whether they are relevant or irrelevant.
- Instead of deciding based on beginning words, we read the whole comment to gain context. Sometimes, the main context will be hidden at the end.
- We thought thoroughly about the context and decided on the appropriate label.
- We identified different and difficult scenarios. We finalized the labels based on a discussion with our group. We ensured that we considered things from everyone's perspective for accurate labels.
- We reviewed and discussed them to provide a consistent dataset to avoid conflicts and disagreements. We spent too much time modifying and improving the dataset, as a consistent and noise-free dataset is key to an accurate model.

- All the annotation guidelines were checked constantly and improved regularly as the team progressed through the annotation process. Regular communication and review helped us get a consistent and high-quality dataset.

Tough Examples:

While annotating we faced certain challenges as we faced some difficult examples to labels.

They were very confusing to give a label.

- Wear gloves - It is one of the steps in process of cooking but that is not having anything related to recipes. Label —IRRELEVANT
- Making Boeuf Bourguignon - It is a different name which is unfamiliar to us, we cannot identify whether it is related to recipes or not until we get what is is —RELEVANT
- Best hot sauce ever- It is opinion regarding the sauce. There is sauce but sauce might be a ingredient or dish name depending on the context. But it is not having any steps regarding sauce —IRRELEVANT
- That's too long for me. It would ruin the effects of reverse searing- If we do not understand what is searing is, we cannot decide it is relevant, searing is process of slow cooking of meat —RELEVANT
- A lot of non Americans don't grow up eating it so the texture was really unfamiliar to me- It is saying about an opinion or experience of American texture, it is style of cooking but not related to recipes as it is just an experience —IRRELEVANT
- Cabbage roles are one of my favourite foods! How do you make them?- It is having name of item but not having any procedure about how to make it. Since there is no process you might wonder it is irrelevant but it is having dish name, recipe has dish, ingredients and procedure —RELEVANT
- I go through so many paper towels whenever I'm in the kitchen. -It is a nice kitchen suggestion to follow but it is not having anything about recipe —IRRELEVANT
- Japanese, ~~Chinese~~***Taiwanese***or Korean -It is having cuisine names but do not have any food item names or ingredients. —IRRELEVANT
- My fave!! Recipe? -It is having recipe name but no contents about recipe —IRRELEVANT

Inter-Annotator Agreement:

The Inter-Annotator Agreement (IAA) was a very crucial aspect of this project, as in this project, we have done annotations in groups by collaborating. When there are multiple annotators, there is scope for clashes due to different thinking and perspectives. They must be solved for running a team towards a common goal. While annotation guidelines will act as common rules and standards to be followed for maintaining unity and getting consistent datasets, IAA will help there by finding room for improvement in the annotated datasets. It quantifies the extent to which different annotators make the same labeling decisions

To practically implement IAA calculations, especially with larger datasets or more than two annotators, there are some software tools or programming libraries designed for analyzing the datasets annotated by several annotators. Python libraries like sklearn for Cohen's Kappa Krippendorff's Alpha are useful resources.

We have used the Chen's Kappa approach for both the datasets: the annotated evaluation dataset and the annotated trained dataset to get the insights of metrics through a Python program. We simply given annotated datasets to the program and got below output:

Annotation Statistics

Attribute	Value	
Examples	800	
Categories	2	
Co-Incident Examples	400	
Single Annotation Examples	0	
Annotators	4	
Avg. Annotations per Example	2.0	
* (>1 annotation)		

Agreement Statistics

Statistic	Value	
Percent (Simple) Agreement	0.47	
Cohen's Kappa (adi vs vish)	0.077	
Cohen's Kappa (adi vs hem)	0.471	
Cohen's Kappa (adi vs jay)	0.357	
Cohen's Kappa (vish vs hem)	0.030	
Cohen's Kappa (vish vs jay)	0.438	
Cohen's Kappa (hem vs jay)	0.189	

** The above output is for evaluation dataset built by four annotators- adi, vish, hem and jay. It is for the evaluation set just made for the without involving the model.

Annotation Statistics

Attribute	Value	
Examples	800	
Categories	2	
Co-Incident Examples	200	
Annotators	4	
Avg. Annotations per Example	4.0	

Agreement Statistics

Statistic	Value	
Cohen's Kappa (adi vs vish)	0.077	
Cohen's Kappa (adi vs hem)	0.471	
Cohen's Kappa (adi vs jay)	0.357	
Cohen's Kappa (vish vs hem)	0.030	

Cohen's Kappa (vish vs jay)	0.438
Cohen's Kappa (hem vs jay)	0.189

** The above output we got when we calculated the IAA for the training and target dataset (800) annotated by annotators. We included IAA agreements for the both datasets evaluation and the labelled dataset with training and target datasets.