HOME WORK-2: TEXT CLASSIFIER MODEL FOR RECIPES
Group 2: EMBEDDING EAGLES

Objective: To determine cuisine from the Reddit post or comment from r/Cooking. The main constraint is it should be a multi-clas classification model. The previous model was multi-label classification model.

Theory: Based on our knowledge, there are three main models in machine learning: Binay Classifier, Multi-label Classifier and Multi-class classification. In Binary classification there will be Two labels only. The label will be first one or another. In Multi-label classification, there can be multiple labels allowed. But in the Multi-class classification, only one label will be allowed out of the multiple labels.
Example:
Binary Classification:- Deciding whether a mail is spam or not. If it is not spam, then it is automatically not spam.
Multi-Label Classification:- Deciding genre of a movie. A movie can be classified into any genre like suspense, horror, thriller, science fiction etc. A movie can be mix of two or three genres too.
Multi-Class Classification:- Deciding animal based on the image. It will have lot of labels like Dog, Cat, Panda etc but it will be one animal at a time. If it is dog then it cannot be a cat or panda.

Approach:

- The basic idea behind the approach was to decide labels first and start annotations with the labels and create one csv file with data from the labelled dataset. Use it to construct the model by splitting the data into two sets training set and target set. First we will train the model with training and evaluate the model stats by passing the target dataset.
- This project does not require creating model and whole coding. We just need to handover the proposal and then understand, start and do some coding to get the bonus points.
- We are planning to take cuisine names as some labels but there are lot of cuisine names to start with. We thought of going on continent wise at first but it does not sound appropriate.
- But we did analyzing certain data entries during annotation process and picked labels of most repeated and famous cuisines like Japanese, Indian, Mexican etc.
- Then we started annotating in prodigy with the labels just like the last project and then we went with exporting the dataset as jsonl. But we wanted to do it using CSV file, so we converted the JSONL dataset into CSV file with the prodigy tool.
- After getting the csv file, we just went doing a part of model. We understood the code, packages and we tried to do implement the model in similar way with the different dataset created by us. We did only small progress as it was for bonus points.
- We included all the coding done by us on our dataset in the Jypter notebook file. We used this video as reference for coding to build the multi-class classification model.for the bonus.
  MULTI CLASS CLASSIFICATION OF TEXT USING MACHINE LEARNING