

## AP1

Cassandra Calciano, Shaily Bajracharya, Kelly Lam

### 2. Answers to the following questions from [Datasheets for Datasets](#)

- A. Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor–patient confidentiality, data that includes the content of individuals’ non-public communications)?**

No, this dataset does not contain data that might be considered confidential.

- B. Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset?**

No personal information can be found directly as all the data in this dataset comes from cookbooks.com website where every recipe just has ingredients and the process for the cooking.

- C. Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)?**

There is no direct sensitive content as it only contains recipes.

- D. If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?**

We will run a simple random sample to select 500 recipes.

- E. Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?**

No, the raw data was not saved.

- F. Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?**

The measurements for recipes use the imperial system. All values are saved as strings, including measurements.

**G. Are there tasks for which the dataset should not be used?**

No, this dataset can be used for any task.

**H. Have any third parties imposed IP-based or other restrictions on the data associated with the instances?**

No.