The datasets are taken from top 2 Indian cooking channel named Nisha Madhulika channel and Kabita’s Kitchen channel.

Both the datasets are divided into seven categories :-

Label 1- Gratitude

Label 2- About the recipe

Label 3- About the video

Label 4- Praising

Label 5- Hybrid

Label 6- Undefined

Label 7- Suggestions and queries

All the labelling has been done manually.

**Nisha Madhulika dataset:**

Dataset characteristics: Multivariate

Number of instances: 4900

Area: Cooking

Attribute characteristics: Real

Number of attributes: 3

Date donated: March, 2019

Associate tasks: Classification

Missing values: Null

Number of subscribers: 7,063,604

**Kabita Kitchen dataset:**

Dataset characteristics: Multivariate

Number of instances: 4900

Area: Cooking

Attribute characteristics: Real

Number of attributes: 3

Date donated: March, 2019

Associate tasks: Classification

Missing values: Null

Number of subscribers: 4,867,502

There are two separate datasets file of each channel. The files with preprocessing names are generated after doing the preprocessing and exploratory data analysis on both the datasets. This file includes:

* Id
* Comment text
* Labels
* Count of stop-words
* Uppercase words
* Hashtags
* Word count
* Char count
* Average words
* Numeric

The main file includes:

* Id
* comment text
* Labels