

Tech Week '22

Machine Learning Hackathon

ACM - IEEE

Introduction

Cooking shows on YouTube are gaining popularity day by day. Whether it's regional, fusion, authentic, we've got channels for everything. But these channels won't be anything without us, the viewers. There are many ways to engage with these content creators from subscribers, likes, and comments. From gratitude to praise, criticism to queries, there are many categories into which we can classify these comments.

Can you build a model that predicts the category of these comments?

Disclaimer: The dataset for this competition contains text that may be considered profane, vulgar, or offensive.

The Challenge

In this competition, you will be categorizing these Hinglish comments into seven classes -

- Label 1 - Appreciation
- Label 2 - Recipe-Specific
- Label 3 - Video-Specific
- Label 4 - Praising
- Label 5 - Mixed
- Label 6 - Miscellaneous
- Label 7 - Suggestions and queries

About the Dataset

The dataset contains 7,840 instances of such comments in Hinglish on which you'll train your model. The hashed Youtube IDs, their comments, and the corresponding labels are also provided.

Drive Link for the Dataset - [Link to Dataset](#)

Evaluation

We will evaluate your submission on these three tasks -

1. Data preprocessing (15%)
2. Exploratory Data Analysis (25%)
3. Performance of your Model (60%)

The final score will comprise weights - 15% for Task-1, 25% for Task-2, 60% for Task-3. Your code will be evaluated against the test dataset provided on the drive.

You are not supposed to shuffle the test dataset, as we'll be evaluating your submissions with the labels in a particular order.

Submission Format

You are required to submit a zip file containing

1. Jupyter notebook/Python File (.ipynb, .py file)
2. A document (.docx/ .pdf) explaining your approach to the problem (**If you do not submit this document, your code will not be evaluated**)
3. A CSV file with the labels for each comment in the test dataset. (Refer to sample submission.csv)

Some pointers to keep in mind -

- Only one person in a team is required to submit. (If you're participating in a team)
- You are not required to comment on your code.
- You can only submit your zip file once. **Any requests for resubmissions will not be entertained.**
- **PLEASE DO NOT SHUFFLE THE TEST DATASET.**

Form Link for Submission - [Form Link for Submission](#)

The competition will be active from 31st Jan 9:00 PM to 4th February 11:59 PM.

In case of any queries, feel free to contact -

Anuradha Pandey - 7045730796

Garvit Kashyap - 9958245103