As YouTube becomes one of the most popular video-sharing platforms, YouTuber is developed as a new type of career in recent decades. YouTubers earn money through advertising revenue from YouTube videos, sponsorships from companies, merchandise sales, and donations from their fans. In order to maintain a stable income, the popularity of videos become the top priority for YouTubers. Meanwhile, some of our friends are YouTubers or channel owners in other video-sharing platforms. This raises our interest in predicting the performance of the video. If creators can have a preliminary prediction and understanding on their videos’ performance, they may adjust their video to gain the most attention from the public.

You have been provided details on videos along with some features as well. Can you accurately predict the number of likes for each video using the set of input variables?

Data Description

**Train Set**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Variable Type** | **Variable Description** |
| video\_id | ID | Identifier for each video |
| title | Text | Name of the Video on Youtube |
| channel\_title | Text/Category | Name of the Channel on Youtube |
| category\_id | Category | Category of the Video (anonymous) |
| publish\_date | Date | The date video was published |
| tags | Text | Different tags for the video |
| views | Numerical | Number of views received by the Video |
| dislikes | Numerical | Number of dislikes on the Video |
| comment\_count | Numerical | Number on comments on the Video |
| description | Text | Textual description of the Video |
| country\_code | Category | Country from which the Video was published |
| likes | Target | (Target) Number of Likes on the video |

**Test Set**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Variable Type** | **Variable Description** |
| video\_id | ID | Identifier for each video |
| title | Text | Name of the Video on Youtube |
| channel\_title | Text/Category | Name of the Channel on Youtube |
| category\_id | Category | Category of the Video |
| publish\_date | Date | The date video was published |
| tags | Text | Different tags for the video |
| views | Numerical | Number of views received by the Video |
| dislikes | Numerical | Number of dislikes on the Video |
| comment\_count | Numerical | Number on comments on the Video |
| description | Text | Textual description of the Video |
| country\_code | Category | Country from which the Video was published |

**Sample Submission:**

Format for making the submission for predictions on the test set

**video\_id:** Unique id for each video in the test set

**likes:** Predicted Number of likes received for that video

## Evaluation Metric

The evaluation metric for this hackathon is 1000\*RMSLE (Root Mean Squared Log Error)

## ****Public and Private split****

The public leaderboard is based on 40% of test data, while final rank would be decided on remaining 60% of test data (which is private leaderboard)

## ****Guidelines for Final Submission****

Please ensure that your final submission includes the following:

1. Solution file containing the prediction of number of likes for videos in the test set
2. Code file for reproducing the submission, note that it is mandatory to submit your code for a valid final submission

**Hackathon Rules**

1. This Hackathon is for learning purposes and there are no AV points as reward
2. Setting the final submission is recommended. Without a final submission, the submission corresponding to best public score will be taken as the final submission
3. Use of external datasets is not allowed and will lead to disqualification from the leaderboard
4. Entries submitted after the contest is closed, will not be considered for leaderboard
5. The code file pertaining to your final submission is mandatory while setting final submission
6. Throughout the hackathon, you are expected to respect fellow hackers and act with high integrity.