**Upvotes Model**

****

**Problem Statement:**

An online question and answer platform has hired you as a data scientist to identify the best question authors on the platform. This identification will bring more insight into increasing the user engagement. Given the tag of the question, number of views received, number of answers, username and reputation of the question author, the problem requires you to predict the upvote count that the question will receive.

## Dataset Description:

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| ID | Question ID |
| Tag | Anonymised tags representing question category |
| Reputation | Reputation score of question author |
| Answers | Number of times question has been answered |
| Username | Anonymised user id of question author |
| Views | Number of times question has been viewed |
| Upvotes | (Target) Number of upvotes for the question |

**Expectations and Evaluation:**

## We expect the candidate to build a model upon training data (70% of the train\_upvotes.csv) and evaluate the model on the validation data, (30% of the train\_upvotes.csv). Use the random\_state=10 while splitting the data. Use the model to predict upon the test\_upvotes.csv.

## Candidates are free to use any of the algorithms to build the model and the evaluation metric for this competition is RMSE on the validation data.