**HR Analytics Problem Statement**

Your client is a large MNC and they have 9 broad verticals across the organization. One of the problems your client is facing is around identifying the right people for promotion and prepare them in time. Currently the process, they are following is:

1. They first identify a set of employees based on recommendations/ past performance
2. Selected employees go through the separate training and evaluation program for each vertical. These programs are based on the required skill of each vertical
3. At the end of the program, based on various factors such as training performance, KPI completion (only employees with KPIs completed greater than 60% are considered) etc., employee gets promotion

For above mentioned process, the final promotions are only announced after the evaluation and this leads to delay in transition to their new roles. Hence, company needs your help in identifying the eligible candidates at a particular checkpoint so that they can expedite the entire promotion cycle.

They have provided multiple attributes around Employee's past and current performance along with demographics. Now, the task is to predict whether a potential promotee at checkpoint in the test set will be promoted or not after the evaluation process.

**Dataset Description**

|  |  |
| --- | --- |
| Variable | Definition |
| employee\_id | Unique ID for employee |
| department | Department of employee |
| region | Region of employment (unordered) |
| education | Education Level |
| gender | Gender of Employee |
| recruitment\_channel | Channel of recruitment for employee |
| no\_of\_trainings | no of other trainings completed in previous year on soft skills, technical skills etc. |
| age | Age of Employee |
| previous\_year\_rating | Employee Rating for the previous year |
| length\_of\_service | Length of service in years |
| KPIs\_met >80% | if Percent of KPIs(Key performance Indicators) >80% then 1 else 0 |
| awards\_won? | if awards won during previous year then 1 else 0 |
| avg\_training\_score | Average score in current training evaluations |
| is\_promoted | (Target) Recommended for promotion |

**Evaluation Metric**

The evaluation metric for this competition is F- Score.

**Predicting Upvotes - 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 |

**Evaluation Metric**

The evaluation metric for this competition is RMSE (root mean squared error)