About Practice Problem: HR Analytics

HR analytics is revolutionising the way human resources departments operate, leading to higher efficiency and better results overall. Human resources has been using analytics for years. However, the collection, processing and analysis of data has been largely manual, and given the nature of human resources dynamics and HR KPIs, the approach has been constraining HR. Therefore, it is surprising that HR departments woke up to the utility of machine learning so late in the game. Here is an opportunity to try predictive analytics in identifying the employees most likely to get promoted.

Data Science Resources

* Refer this [comprehensive guide](https://www.analyticsvidhya.com/blog/2017/03/imbalanced-classification-problem/) that will help you to understand how to tackle classification problems.
* Also, check out the [solutions](https://www.analyticsvidhya.com/blog/2018/12/wns-hackathon-solutions-by-top-finishers/) provided by the folks who seured top ranks in this hackathon
* Are you a complete beginner? if yes, you can check out our latest ['Introduction to Data Science'](https://trainings.analyticsvidhya.com/courses/course-v1:AnalyticsVidhya+DS101+2018T2/about?utm_source=practice_problem_HR_Analytics&utm_medium=Datahack) course to kickstart your journey in data science.

Rules

* One person cannot participate with more than one user accounts.
* This is proprietary dataset, you can only use for this hackathon (Analytics Vidhya Datahack Platform) not for any other reuse
* You are free to use any tool and machine you have rightful access to.
* You can use any programming language or statistical software.
* You are free to use solution checker as many times as you want.

FAQs

**1. Are there any prizes/AV Points for this contest?**

This contest is purely for learning and practicing purpose and hence no participant is eligible for prize or AV points.

**2. Can I share my approach/code?**

Absolutely. You are encouraged to share your approach and code file with the community. There is even a facility at the leaderboard to share the link to your code/solution description.

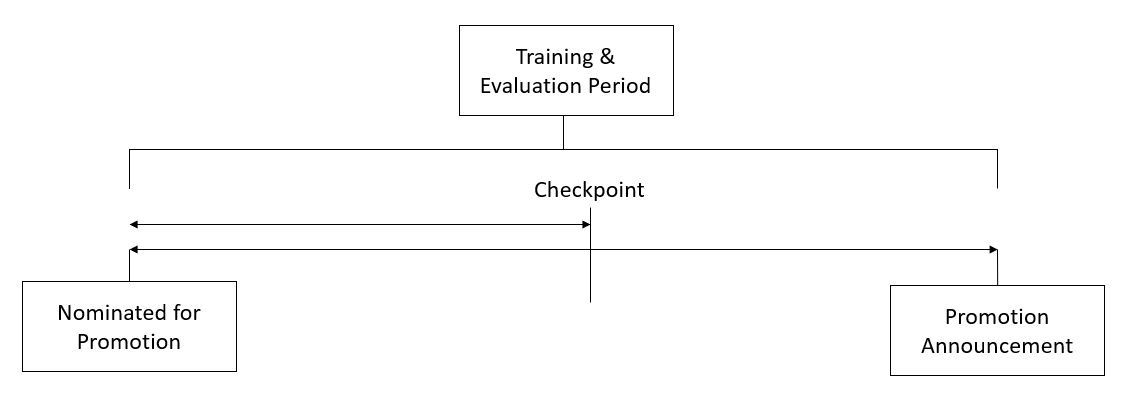
**3. I am facing a technical issue with the platform/have a doubt regarding the problem statement. Where can I get support?**

Post your query on discussion forum at the thread for this problem, discussion threads are given at the bottom of this page. You could also join the AV slack channel by clicking on 'Join Slack Live Chat' button and ask your query at channel: practice\_problems.

**Problem Statement**

Your client is a large MNC and they have 9 broad verticals across the organisation. One of the problem your client is facing is around identifying the right people for promotion (only for manager position and below) 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 F1 Score.

## Public and Private Split

Test data is further randomly divided into Public (40%) and Private (60%) data.

* Your initial responses will be checked and scored on the Public data.
* The final rankings would be based on your private score which will be published once the competition is over.