Technical Board
IIT Guwahati

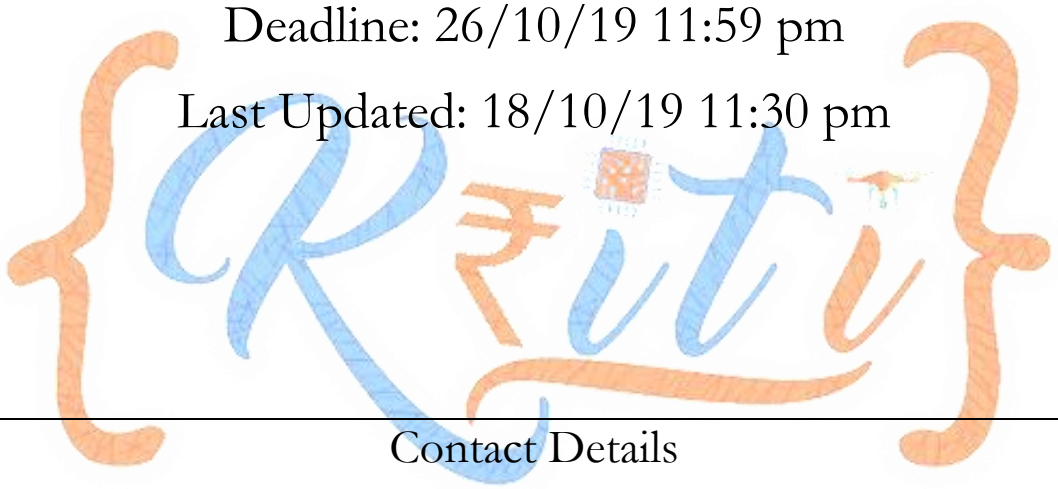
CONSULTING & ANALYTICS CLUB
IIT GUWAHATI

Data Science Hackathon

GC Points: 300

Deadline: 26/10/19 11:59 pm

Last Updated: 18/10/19 11:30 pm



Contact Details	
Rishabh Agrawal	**Krish Tikmani**
Secretary	Analytics Head
9820968371	7903732928

Email: caciitg@gmail.com

Indian Engineering Graduates Employability Rate

1. Problem Statement

"More than 75% of the Engineering graduates in India are not fit to be employed."

The government of India needs to prioritize higher education and undertake long-term policy interventions in the next 5-10 years to ameliorate the low rate of engineering employability.

To understand the reasons for such low employability rates, XYZEE, an Analytics company has surveyed a sample of Engineering graduates about their performance history. The company wants to deploy an application for the Engineering graduates to predict what would be their expected salary after Graduation.

You, as a Data Scientist have to prepare a predictive model for XYZEE and predict the salary of an Engineer based on some parameters defining their past performance.

2. The Competition

To make the hackathon more interesting, we are hosting an online competition at Kaggle to get the best of solutions. So, wear on your "data hats" and join us at:

kaggle https://www.kaggle.com/c/kriti2019cna

3. About the Dataset

a) You are given an anonymized training dataset training_data.csv of **3000** engineering students of India consisting of **35** features (incl. Salary). You have to predict the salary of engineers on an unseen test_dataset.csv based on the remaining 34 features.

b) The description of the features can be found in the data_dictionary.xlsx file provided on the competition site.

c) The submission csv file should contain the headers ['ID', 'Salary']. The csv file needs to be uploaded on the competition site. You can also find the sample submission sample.csv on the competition site.

d) A portion of IDs in the test_data.csv is used to evaluate a Public score while will be used to evaluate Private score.

e) You will be able to see the Public score as soon as you upload a submission on the competition site.

4. Evaluation Criteria:

i. Public Leaderboard Score: 25%

ii. Private Leaderboard Score: 75%

5. Evaluation Metrics:

R-Squared (Coefficient of Determination)

6. The Rules of the Game

The participants must strictly adhere to the rules given below. Violating any of the rules will lead to disqualification of the team and/or penalty to the hostel.

- The team name format should be: 'teamName_hostelName'. For example: '**"TeamA1_Lohit".** Any team not following the format will be **not** considered for evaluation.

- **Any number of teams can participate from every hostel.** However, in the end only the top scoring team from each hostel will be considered for the final rankings between Hostels.

- Total score: **25% Public Leaderboard + 75% Private Leaderboard**

- The top performers from each hostel will have to submit their notebooks/scripts at caciitg@gmail.com.

- Plagiarism/Cheating will lead to disqualification. If the top performer from a hostel is found to be violating any rule, the **next performing team** from the given hostel will be considered for final evaluation.

- Final Hostel Rankings/ GC points will be relative and will be declared at the end of the competition.