



(/teams/codefest-enigma/create)



(/contest/codefest-enigma/lb)

**codeFest.<sup>17</sup>**  
22nd - 24th September

**Analytics Vidhya**  
Learn everything about analytics

**ENIGMA**  
Machine Learning

22ND SEPTEMBER  
08:00 PM IST  
TO  
24TH SEPTEMBER  
08:00 AM IST

PRIZES WORTH  
**₹ 50,000**

Registered

<b>Starts at</b>	Fri Sep 22 2017 20:00:00 GMT+0530 (IST)
<b>Closes on</b>	Sun Sep 24 2017 08:00:00 GMT+0530 (IST)
<b>Mode</b>	Online
<b>Fee</b>	Free
<b># Participants</b>	416
<b>Prizes</b>	Prizes worth INR 50K and top 20 will get discounts on DataHack Summit tickets

**Contest ends in 1 days: 10 hours: 46 minutes: 59 seconds**

Hackathons are becoming a popular way for individuals to showcase their machine learning expertise and knowledge. Hackathon mania is a new fever amongst data scientists and machine learning enthusiasts.

Come be part of this competition that you will give you more reasons to love hackathons.

## Prizes:

Top 10 rankers of the hackathon will be awarded

- Rank 1: INR 20K
- Rank 2: INR 12K
- Rank 3: INR 8K
- Rank 4-10: INR 1K
- Rank 1 in IIT (BHU) Varanasi: INR 3K

This is not all, if you get in the top 20, you will get exclusive benefits on India's Largest Data Science Event, in Bengaluru, DataHack Summit 2017 (<https://www.analyticsvidhya.com/datahacksummit/>)

- Top 10 contestants will get a 25% OFF on Conference Tickets (<https://www.analyticsvidhya.com/datahacksummit/all-tickets/>) (2 days and 3 days Pass)
- Contestants from ranks 11-20 will get a 15% OFF on Conference Tickets (<https://www.analyticsvidhya.com/datahacksummit/all-tickets/>) (2 days and 3 days Pass)

## Rules

### Contest Guidelines

- One person cannot participate with more than one user accounts.
- Appropriate taxes will be applicable on the prize money.
- You cannot use ID as a feature for modelling
- You cannot use future data for modelling
- You should submit a .zip for submission

## Tools



- You are free to use any tool and machine you have rightful access to.
- You can use any programming language or statistical software.

## Solution Checker

- You are free to use solution checker as many times as you want.
- Adding comment is mandatory for use of solution checker
- Comments will help you to refer to a particular solution at a later point in time.

## Final Submissions

- Setting final submission is mandatory. If you don't make final submission, your entry would be disqualified.
- No submissions would be entertained after the hackathon ends.
- Code file is mandatory while setting final submission. For GUI based tools, upload zip file of snapshots of steps taken by you, else upload code file.
- The code file uploaded should be pertaining to your final submission. If we find any discrepancy between the two, your entry would be dis-qualified.

## Team formation

- Maximum of 2 people can form a team.
- In case a team wins, prize would be distributed equally among team members
- Team once created can't be dissolved.
- Teams can't be merged.

## Expected conduct

- At any point in the hackathon, you are expected to respect fellow hackers and act with high integrity.
- Slack Live Chat admins hold the right to blacklist / block any participant found to use foul / disrespectful language. Chat forum will be closely monitored.
- Analytics Vidhya holds the right to disqualify any participant at any stage of competition if found indulged in fraudulent practices.

## Registration Fee

Free

## Problem Statement

Online judges provide a platform where many users solve problems everyday to improve their programming skills. The users can be beginners or experts in competitive programming. Some users might be good at solving specific category of problems(e.g. Greedy, Graph algorithms, Dynamic Programming etc.) while others may be beginners in the same. There can be patterns to everything, and the goal of the machine learning would be to identify these patterns and model user's behaviour from these patterns.

The goal of this challenge is to predict **range of attempts** a user will make to solve a given problem given user and problem details.

## Motivation

Finding these patterns can help the programming committee, as it will help them to suggest relevant problems to solve and provide hints automatically on which users can get stuck.

## Data

We have collected user submissions for various problems from an online judge. The submissions data consists of 2,21,850 submissions of 3,571 users and 6,544 problems.



## Data Files

There are 3 training data files.

1. **train\_submissions.csv** - This contains 1,55,295 submissions which are selected randomly from 2,21,850 submissions. Contains 3 columns ('user\_id', 'problem\_id', 'attempts\_range'). The variable 'attempts\_range' denoted the range no. in which attempts the user made to get the solution accepted lies.

We have used following criteria to define the attempts\_range :-

attempts_range	No. of attempts lies inside
1	1-1
2	2-3
3	4-5
4	6-7
5	8-9
6	>=10

1. **user\_data.csv** - This is the file containing data of users. It contains the following features :-

1. user\_id - unique ID assigned to each user
2. submission\_count - total number of user submissions
3. problem\_solved - total number of accepted user submissions
4. contribution - user contribution to the judge
5. country - location of user
6. follower\_count - amount of users who have this user in followers
7. last\_online\_time\_seconds - time when user was last seen online
8. max\_rating - maximum rating of user
9. rating - rating of user
10. rank - can be one of 'beginner', 'intermediate', 'advanced', 'expert'
11. registration\_time\_seconds - time when user was registered

1. **problem\_data.csv** - This is the file containing data of the problems. It contains the following features :-

1. problem\_id - unique ID assigned to each problem
2. level\_id - the difficulty level of the problem between 'A' to 'N'
3. points - amount of points for the problem
4. tags - problem tag(s) like greedy, graphs, DFS etc.

1. **test\_submissions.csv** - This contains the remaining 66,555 submissions from total 2,21,850 submissions. Contains 1 column (ID). The 'attempts\_range' column is to be predicted.

## Submission Details

The following 3 files are to be uploaded.

1. **test\_predictions.csv** - This contains the predictions of the model on the test\_submissions.csv file. The file must contain 2 columns - (ID, 'attempts\_range'). (where ID is the concatenation of user\_id and problem\_id). Please ensure that the order of the predictions in this file matches the order of the input ratings in test\_submissions.csv.

1. **A PDF file detailing the model** - Explain the preprocessing steps and the approach (along with proper justification for choosing a specific model or strategy). Observations and insights derived from the various efforts put into this problem are welcome as well.

1. **A .zip file of source code** - The code should produce the output file submitted and must be properly commented.



## Evaluation Metric:

The metrics used for evaluating the performance of the model is the F1 score between the predicted and the actual value with average parameter set to 'weighted' .

📄 Test File (/contest/codefest-enigma/download/test-file)

📄 Train File (/contest/codefest-enigma/download/train-file)

📄 Sample Submissions (/contest/codefest-enigma/download/sample-submission)

---

## Solution Checker

### Code File

Choose File No file chosen

### Solution File\* (.csv, .zip only)

Choose File No file chosen

### Solution Description (max : 180 chars)\*

Solution Description (for your interest)

Add Solution

Join Slack Live Chat






## CONFERENCE ON ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

9-11 November, Bengaluru



Keynote Speaker  
Dr. Kirk Borne

**Buy Tickets**

(<https://www.analyticsvidhya.com/datahacksummit/>)



#### ANALYTICS VIDHYA

About Us

(<http://www.analyticsvidhya.com/about-me/>)

Our Team

(<https://www.analyticsvidhya.com/about-me/team/>)

Career

(<https://www.analyticsvidhya.com/career-analytics-vidhya/>)

Contact Us

(<https://www.analyticsvidhya.com/contact/>)

#### DATA SCIENTISTS

Blog

(<https://www.analyticsvidhya.com/blog/>)

Hackathon

(<https://datahack.analyticsvidhya.com/>)

Discussions

(<https://discuss.analyticsvidhya.com/>)

Apply Jobs

(<https://www.analyticsvidhya.com/jobs/>)

Leaderboard

(<https://datahack.analyticsvidhya.com/leaderboard/>)

#### COMPANIES

Post Jobs

(<https://www.analyticsvidhya.com/jobs/company/signup/>)

Trainings

(<https://www.analyticsvidhya.com/trainings/>)

Hiring Hackathons

(<https://www.analyticsvidhya.com/hiring-hackathons/>)

Advertising

(<https://www.analyticsvidhya.com/contact/>)

Reach Us

(<https://www.analyticsvidhya.com/contact/>)

#### JOIN OUR COMMUNITY :

f

(<http://www.facebook.com/Analyticsvidhya>)

38665

(<http://www.facebook.com/Analyticsvidhya>)

Followers

(<http://www.facebook.com/Analyticsvidhya>)

11260

(<http://www.facebook.com/Analyticsvidhya>)

Followers

(<https://plus.google.com/+Analyticsvidhya>)

2192

(<https://plus.google.com/+Analyticsvidhya>)

Followers

(<https://www.linkedin.com/company/analyticsvidhya>)

beta/3632032/

Subscribe to emailer



© Copyright 2013-2017 Analytics Vidhya.

Privacy Policy (<https://www.analyticsvidhya.com/privacy-policy/>)

Terms of Use (<https://www.analyticsvidhya.com/terms/>)

Refund Policy (<https://www.analyticsvidhya.com/refund-policy/>)

