hackerrank.com

HackerRank

original

Quora Machine Learning CodeSprint

7029 views, including:

Robert Johnson
Via Software Engineering

Johnny Ho, random.johnnyh
Via the blog Engineering at Quora

Grace Gallogly, Quora User Research Intern
Via the blog Engineering at Quora

Oliver Emberton, founder of Silktide
Via the blog Engineering at Quora

Garindra Prahandono, Software Engineering Intern @ Sony USA
Via Kah Seng Tay

Ng Boon Leong, Maths and Physics in college, former maths teacher
Via Serene Lee

Nikhil Garg, Dreamer, Programmer, Blogger
Via the blog Engineering at Quora

Quora Machine Learning CodeSprint

7029 views, including:

Robert Johnson
Via Software Engineering

Johnny Ho, random.johnnyh
Via the blog Engineering at Quora

Grace Gallogly, Quora User Research Intern
Via the blog Engineering at Quora

Oliver Emberton, founder of Silktide
Via the blog Engineering at Quora

Garindra Prahandono, Software Engineering Intern @ Sony USA
Via Kah Seng Tay

Ng Boon Leong, Maths and Physics in college, former maths teacher
Via Serene Lee

Nikhil Garg, Dreamer, Programmer, Blogger
Via the blog Engineering at Quora

Quora uses a combination of machine learning algorithms and moderation to ensure high-quality content on the site. High question and answer quality has helped Quora distinguish itself from other Q&A sites on the web.

Not all of these questions are as interesting or appealing to people on the web. Can you tell what questions can organically attract the most viewers? What about questions that eventually become viral? Which questions are timeless and can sustain traffic?

For this task, given Quora question text, topic data, number of answers and number of people promoted to, predict the number of views per day in age of the question.

Input Format

The first line contains N. N questions follow, each being a valid json object. The following fields of raw data are given in json.

- question_key (string): Unique identifier for the question.
- question_text (string): Text of the question.
- context_topic (object): The primary topic of a question, if present. Null otherwise. The topic object will contain a name (string) and followers (integer) count.
- topics (array of objects): All topics on a question, including the primary topic. Each topic object will contain a name (string) and followers (integer) count.
- anonymous (boolean): Whether the question was anonymous.
- num_answers (integer): The number of visible non-collapsed answers the question has.
- promoted_to (integer): The number of people the question was promoted to.
- __ans__ (float): The ratio of viewers to age of the question in days.

This is immediately followed by an integer T.
T questions follow, each being a valid json object.
The json contains all but one field *__ans__*.

Output Format

T rows of JSON encoded fields, with the *question_key* key containing the unique identifier given in the test data, and the predicted value keyed by *__ans__*.

Constraints

question_key is of ascii format.
question_text, name in topics and context_topic is of UTF-8 format.
0 <= followers <= 10^6
9000 <= N <= 45000
1000 <= T <= 5000

Sample Input

```
 9000
json_object
json_object
....
json_object
1000
json_object
json_object
....
json_object
```

Sample Output

```
 json_object
json_object
...
json_object
```

Sample testcases can be downloaded here and used for offline training if desired.

Scoring

Your solution is evaluated by a Root Mean Squared Logarithmic Error (RMSLE) metric. We then calculate a final score by how close it reaches a target score of 0.5, and scale that by 100%. You can score higher than 100%.

$$\frac{0.5}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log(x_i)-\log(y_i))^2}} \times 100\%$$

where x_i is the predicted value of i^{th} input + 1 and y_i is the actual value of the i^{th} input + 1.

Your score will be based only on the hidden input. The sample input is only for your convenience.

Original URL:
https://www.hackerrank.com/contests/quora/challenges/quora-ml-views