

kaggle

Search



Competitions

Datasets

Kernels

Discussion

Learn



Featured Prediction Competition

Quora Question Pairs

Can you identify question pairs that have the same intent?

\$25,000

Prize Money



Quora · 3,307 teams · 2 years ago

Overview

Data

Kernels

Discussion

Leaderboard

Rules

Late Submission

Overview

Description

Where else but [Quora](#) can a physicist help a chef with a math problem and get cooking tips in return?

Evaluation

Quora is a place to gain and share knowledge—about anything. It's a platform to ask questions and connect with people who contribute unique insights and quality answers. This empowers people to learn from each other and to better understand the world.

Prizes

Timeline

Over 100 million people visit Quora every month, so it's no surprise that many people ask similarly worded questions. Multiple questions with the same intent can cause seekers to spend more time finding the best answer to their question, and make writers feel they need to answer multiple versions of the same question. Quora values canonical questions because they provide a better experience to active seekers and writers, and offer more value to both of these groups in the long term.

Currently, Quora uses a Random Forest model to identify duplicate questions. In this competition, Kagglers are challenged to tackle this natural language processing problem by applying advanced techniques to classify whether question pairs are duplicates or not. Doing so will make it easier to find high quality answers to questions resulting in an improved experience for Quora writers, seekers, and readers.

Leaderboard



- 1 DL guys
- 2 Depp Learning
- 3 Jared Turkewitz & sjv
- 4 YesOfCourse
- 5 Qingchen | KazAnova | Faron
- 6 LAMAA power
- 7 aphex34

Kernels



Data Analysis & XGBoost Start...

1009 votes · 2 years ago

Quora EDA & Model selection (...)

227 votes · 2 years ago

Spell Checker using Word2vec

147 votes · 2 years ago

Magic Features (0.03 gain)

148 votes · 2 years ago

291 discussion topics



Contrastive loss - how to make ...

4 replies · a month ago

1st place solution

63 replies · a month ago

24th Place Solution Repo

17 replies · 2 months ago

predict no_duplicated question...

0 replies · 2 months ago

8 NLPFakers

Simple Leaky Exploration Note...

146 votes · 2 years ago

Test data

3 replies · 2 months ago

Launch
2 years ago



Close
2 years ago

2 years ago

Rules Acceptance
Deadline

3,307 Teams
3,944 Competitors
53,927 Entries

Points This competition awarded standard ranking points
Tiers This competition counted towards tiers

Tags

- internet
- linguistics
- text data
- tabular data
- duplicate detection
- logloss
- small