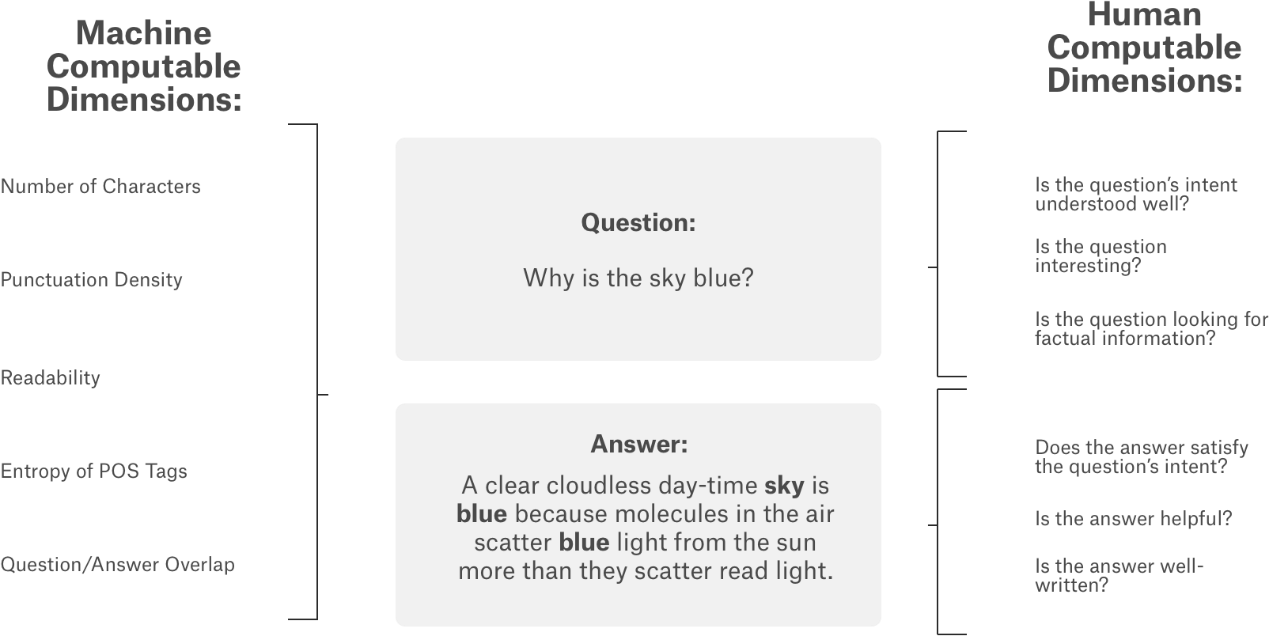
Google QUEST Q&A Labeling

**Description**

Computers are really good at answering questions with single, verifiable answers. But, humans are often still better at answering questions about opinions, recommendations, or personal experiences.

Humans are better at addressing subjective questions that require a deeper, multidimensional understanding of context - something computers aren't trained to do well…yet.. Questions can take many forms - some have multi-sentence elaborations, others may be simple curiosity or a fully developed problem. They can have multiple intents, or seek advice and opinions. Some may be helpful and others interesting. Some are simple right or wrong.



Unfortunately, it’s hard to build better subjective question-answering algorithms because of a lack of data and predictive models. That’s why the [CrowdSource](https://crowdsource.google.com/" \t "_blank)（众包） team at Google Research, a group dedicated to advancing NLP and other types of ML science via crowdsourcing, has collected data on a number of these quality scoring aspects.

In this competition, you’re challenged to use this new dataset to build predictive algorithms for different subjective aspects of question-answering. The question-answer pairs were gathered from nearly **70 different websites**, in a "common-sense" fashion. Our raters(评估者) received minimal guidance and training, and relied largely on their **subjective interpretation of the prompts**. As such, each prompt was crafted in the most intuitive fashion so that raters could simply use their common-sense to complete the task. By lessening our dependency on complicated and opaque rating guidelines, we hope to increase the reuse value of this data set. **What you see is what you get**!

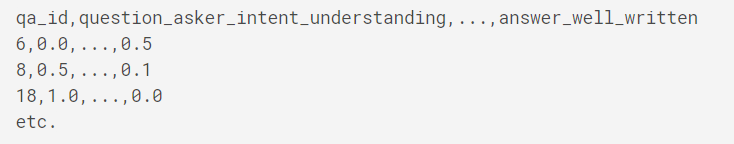
Demonstrating these subjective labels can be predicted reliably can shine a new light on this research area. Results from this competition will inform the way future intelligent Q&A systems will get built, hopefully contributing to them becoming more human-like.

**Evaluation**

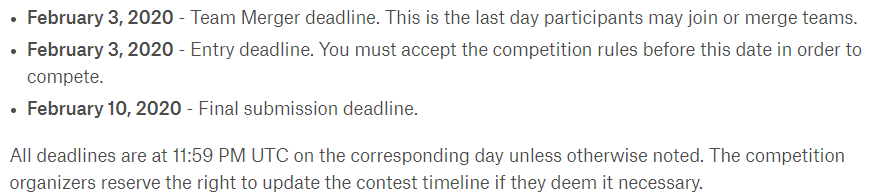
Submissions are evaluated on the mean column-wise [**Spearman's correlation coefficient**](https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient). The Spearman's rank correlation is computed for each target column, and the mean of these values is calculated for the submission score.

**Submission File**

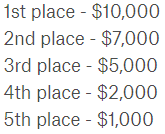
For each qa\_id in the test set, you must predict a probability for each target variable. The predictions should be in the range [0,1]. The file should contain a header and have the following format:



**Timeline**



**Prizes**



**This is a Notebook-only competition**

Submissions to this competition must be made through Notebooks. Your Notebook will re-run automatically against an unseen test set, and needs to output a file named submission.csv. *You can still train a model offline, upload it as a dataset, and use the notebook exclusively to perform inference.*

In order for the "Submit to Competition" button to be active after a commit, the following conditions must be met:

* CPU Notebooks <= 9 hours run-time
* GPU Notebooks <= 2 hours run-time
* Internet must be turned off

In synchronous Notebooks-only competitions, note that a submission that results in an error--either within the notebook or within the process of scoring--will count against your daily submission limit and will not return the specific error message. This is to limit probing the private test set.

Please note - committing your Notebook will result in a full execution of your code from top-to-bottom.

Please see the [Code Competition FAQ](https://www.kaggle.com/docs/competitions#kernels-only-FAQ) for more information on how to submit.

**Data Description**

The data for this competition includes questions and answers from various StackExchange properties. Your task is to predict target values of 30 labels for each question-answer pair.

The list of 30 target labels are the same as the column names in the sample\_submission.csv file. Target labels with the prefix question\_ relate to the question\_title and/or question\_body features in the data. Target labels with the prefix answer\_relate to the answer feature.

Each row contains a single question and a single answer to that question, along with additional features. The training data contains rows with some duplicated questions (but with different answers). The test data does not contain any duplicated questions.

This is not a binary prediction challenge. Target labels are aggregated from multiple raters, and can have continuous values in the range [0,1]. Therefore, predictions must also be in that range.

Since this is a synchronous re-run competition, you only have access to the Public test set. For planning purposes, the re-run test set is no larger than 10,000 rows, and less than 8 Mb uncompressed.

Additional information about the labels and collection method will be provided by the competition sponsor in the forum.

**File descriptions**

* **train.csv** - the training data (target labels are the last 30 columns)
* **test.csv** - the test set (you must predict 30 labels for each test set row)
* **sample\_submission.csv** - a sample submission file in the correct format; column names are the 30 target labels