|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Topic | Kaggle Challenge | Level of interest | Link | Data Size |
| QA label | Google QUEST Q&A Labels | High | https://www.kaggle.com/c/google-quest-challenge/overview | 13Mb |
| Sentimental analysis | Jigsaw Unintended Bias in Toxicity Classification | Medium | https://www.kaggle.com/c/jigsaw-unintended-bias-in-toxicity-classification | 2GB |
| QA label | TREC-COVID Information Retrieval | High | https://www.kaggle.com/c/trec-covid-information-retrieval/data?select=CORD-19 | 20GB |
| Classification, no labelled training | Gendered Pronoun Resolution | Medium | https://www.kaggle.com/c/gendered-pronoun-resolution/data | 7Mb |
| Text parsing | Data Science for Good: City of Los Angeles | Low (Not suitable for project) | https://www.kaggle.com/c/data-science-for-good-city-of-los-angeles/discussion/102388 | 37Mb |