

Introduction

Data is important in supervised classification tasks. However, obtaining the quality data from a diversified population at scale can prove to be a difficult task. MTurk offers an interesting solution as we are able to gather data from a set of anonymous users with different backgrounds. However, there are few challenges of utilizing this approach. We need to ensure that the quality of the data is high from a seemingly random non-expert population. Also, since most users could not be equipped with the database knowledge, we would need to design the interface in a manner that would provide the know-how in a short span of time, thus maximizing the data collection efficiency.

Crowdsourcing Application

We propose a three-tiered design for the task. Namely the tutorial, chat and data collection and review parts. First, the Mturk User is introduced to a tutorial phase. The tutorial website serves as both guiding the user as well as filters out users who are incapable of performing the task. We monitor the user capabilities on the task in the tutorial phase.

An expert is assigned to the user post-tutorial. The chat application mirrors the tutorial interface to provide a seamless experience to the user. The system then assigns a random question from the task bank to the user. Once the task is completed, we assign a score to the task. The Mturk user is remunerated for his efforts and may choose to proceed to the next task or exit the application.

The final part of the system is the review application. A set of reviewers can choose to accept/reject tasks, as well as amend the data applied.

The application is built to be scalable and fault-tolerant. We are able to hold 30 concurrent users at peak without any noticeable slowdowns as well as having an uptime of the system of 99.1%.

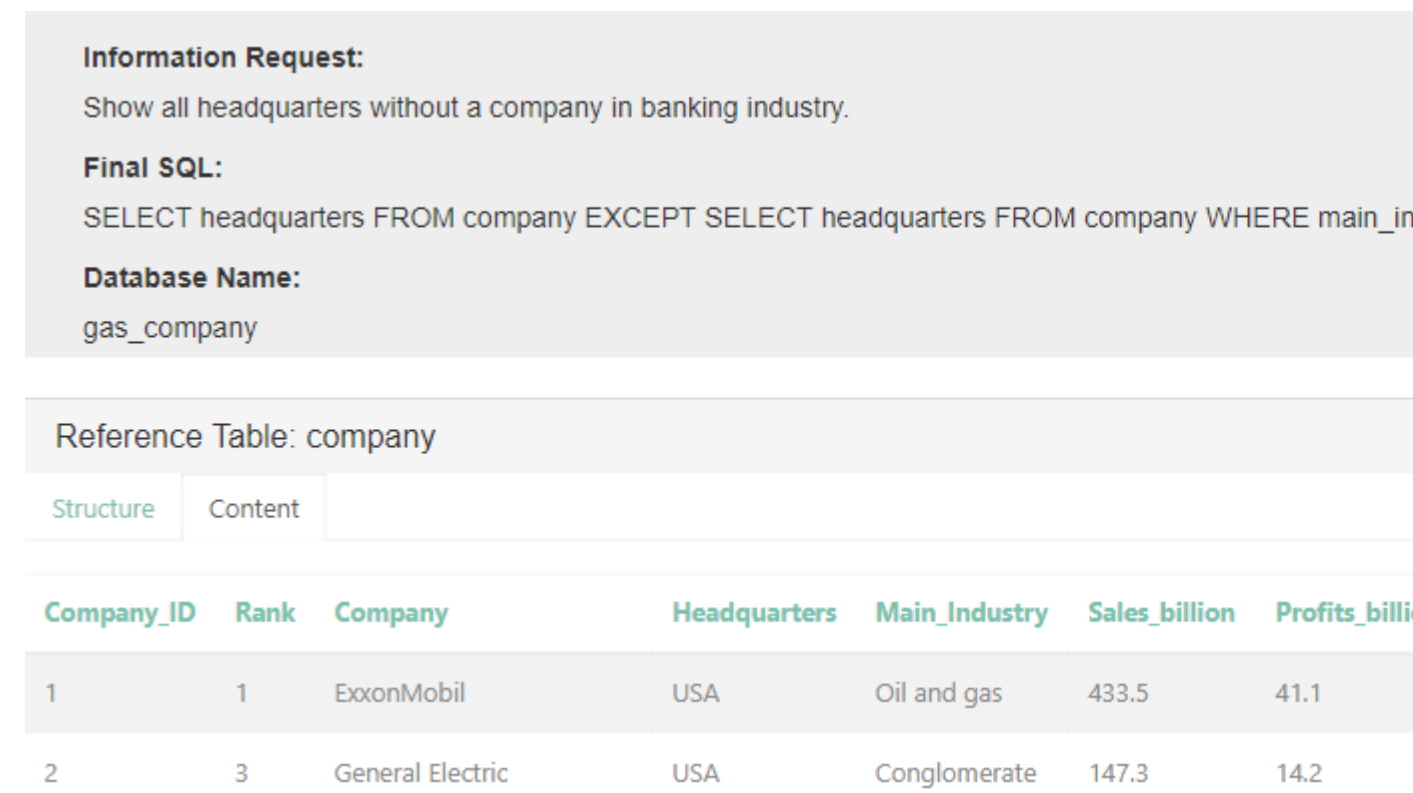


Figure 1. Chat application

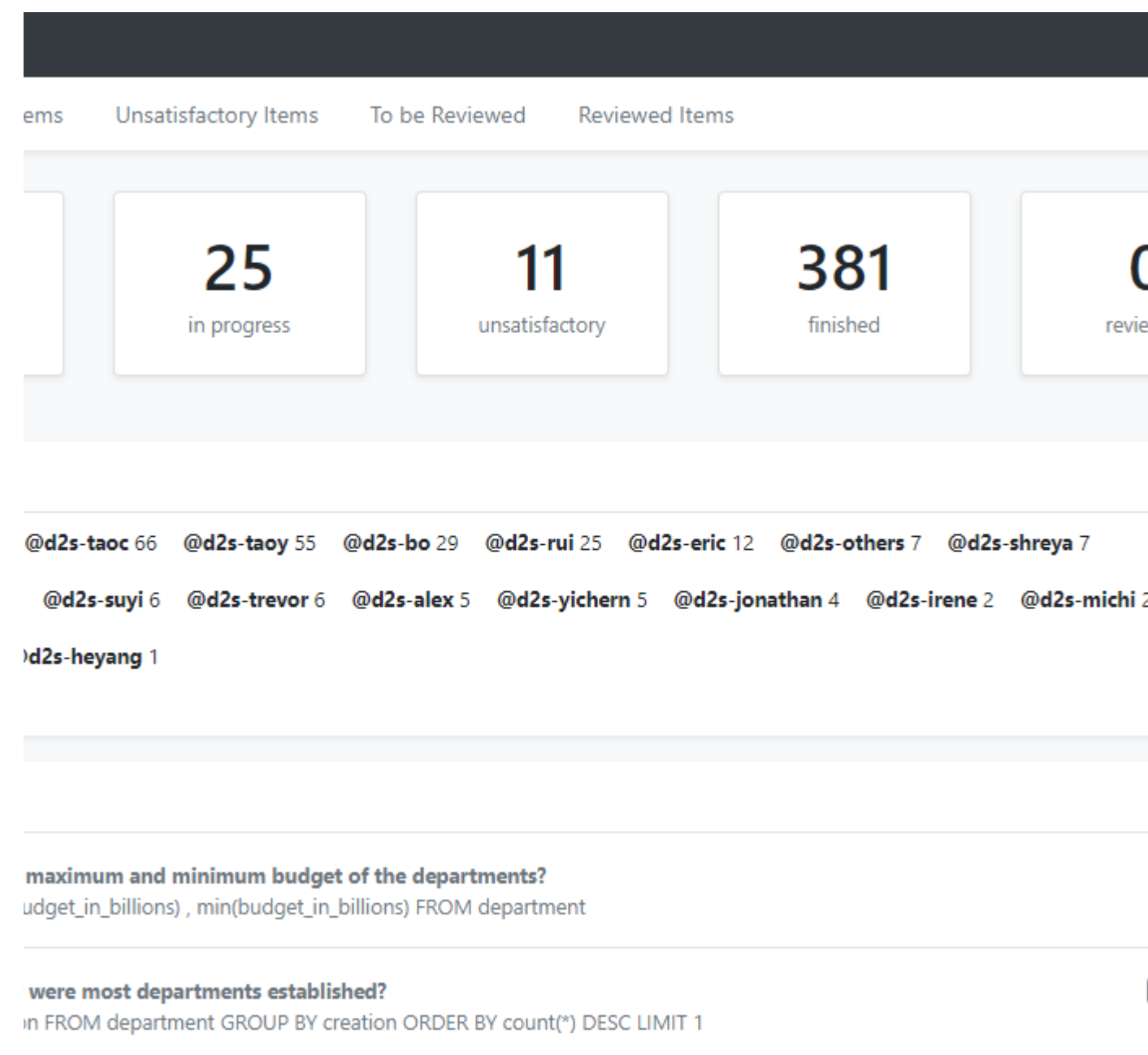


Figure 2. Review application

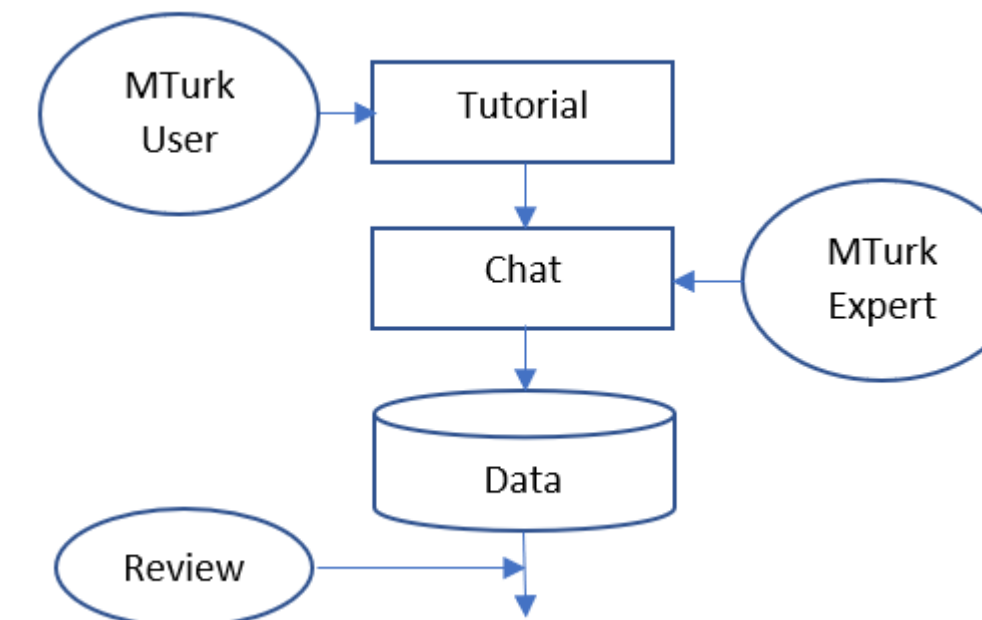


Figure 3: Application Flow

- You are playing the role of a user who wants to know the answer to
 - Ask at least 3 related questions that must include the answer to the question
 - You can break the given question down into small questions, but you are highly encouraged to ask more than 5 good r
 - You'll get \$ 0.5 bonus later if you follow the rules and ask more than 5 good r

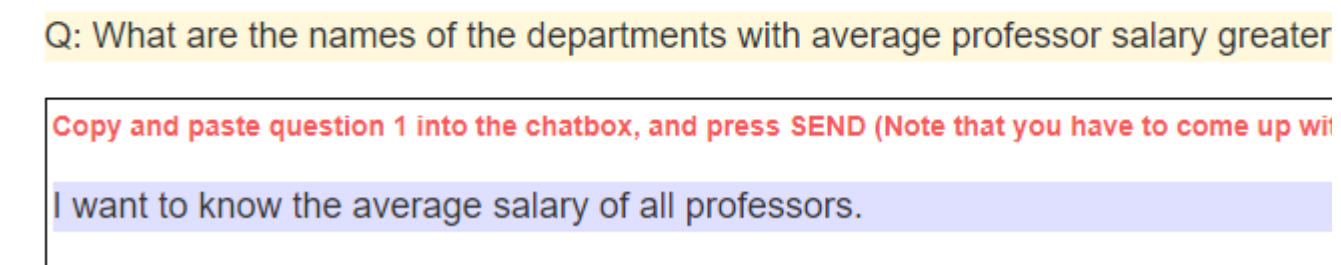


Figure 4: Tutorial Application

Table 1. Crowdsourcing Application Improvement

Version	Failed	4-Rated	5-Rated	Total Accepted
Previous	80%	15%	5%	20%
Current	25%	47.50%	27.50%	75%

Table 2. Crowdsourcing Application: User statistics

	SQL/DB	Excel/Tableau	Task Confidence
Mean Competency (0-5)	2.91	3.66	4.07

Results

We've run the application and obtained 381 sets of completed data in the span of 2 weeks. We provide each Mturk users \$1.50 for a task completion and an additional \$0.50 bonus for a task completion.

For crowdsourcing dialog on querying databases. We noted a 75 percentage of accepted data. This is up 55% from the 20% from the legacy application. Of which, 27.50% of the tasks were rated 5 out of 5 while 47.50% were 4 rated. Of the remaining 25%, the users either left the task midway, or the data provided was unaccepted due to the following reasons: 1) unrelated queries 2) independent queries.

Of the users participating, We record a SQL/Database competency level of 2.91, excel competency of 3.66 and a task confidence of 4.07. The tutorial has a 22.63% pass rate.

Of all the accepted tasks, 46.98% of the users obtains a score of 3, 23.10% of the users has a score of 4 while the rest of the 30% of the users has a score of 5. The average score the data obtained ins 3.83.

Conclusion

We have shown the use of Amazon Mturk in crowdsourcing dialogues. We show a three-stage application which aims to maximize data collection efficiency as well as obtaining high quality data from the user. In our new layout, we show a 375% increase in task accepted with a high meandata score of 3.83.

Acknowledgement

I would like to acknowledge Tao Yu, Prof. Dragomir Radev for his supervision of my project.