**Problem Statement**

Based on the given attributes for Stack overflow questions and answers, build a robust machine learning model that predicts the probability of the question will be answered within 7 days.

**Data**

The dataset is based on stack overflow Questions and Answers across the internet. The rows of Question dataset have rich amount of information regarding question creation date, body, answers count, comments count, scores, tags etc. The Answer dataset contains the information for the answers, answer scores and tags, etc. One Question Id (Parent Id in answer dataset) may have multiple answers rows depending on number of answers.

The questions and answers data comprise of 17,278,709 samples and of 26,496,612 samples each. This dataset has been collected over long time for questions posted since July 2008 to March 2019.

The Questions and Answers tables contain the attributes as mentioned below:

**Questions dataset features**

1. Id (question id)
2. title
3. body (question description)
4. accepted\_answer\_id
5. answer\_count
6. comment\_count (comment count for each question)
7. community\_owned\_date
8. creation\_date (Question Creation date)
9. favorite\_count
10. last\_activity\_date
11. last\_edit\_date
12. last\_editor\_display\_name
13. last\_editor\_user\_id
14. owner\_display\_name
15. owner\_user\_id
16. post\_type\_id
17. score (Upvotes for question – downvotes for question)
18. tags (tags related to question)
19. view\_count (no. of views for the question)

**Answers dataset features**

1. id (Answer id)
2. body (Answer in detail)
3. comment\_count (comment count for each answer)
4. community\_owned\_date
5. creation\_date (answer created date)
6. last\_activity\_date
7. last\_edit\_date
8. last\_editor\_display\_name
9. last\_editor\_user\_id
10. owner\_display\_name
11. owner\_user\_id
12. parent\_id (This belongs to the “Id” from question dataset)
13. post\_type\_id
14. score (Upvotes for answer – downvotes for answer)
15. tags (tags for the answer)

**Please follow the below steps**

1. Open the new incognito window ( Ctrl + Shift + n) and sign in with your Gmail id
2. Go to link - <https://cloudapps.countants.com/>
3. Go to "Sign in" -> Storage Link -> Click ✔ Agree the terms for service and continue
4. Go back to Welcome page again -> Click on "BigQuery Link"
5. Go to Select a project -> Powered by Countants -> Done -> Page will be directed to BigQuery
6. Click on "Add data" (left side of panel)-> Click on “Explore public dataset” -> Search for "Stackoverflow" -> Click on “Stackoverflow” -> Click on “View dataset”
7. Go to your Gmail inbox and open the python notebook (mltest.ipnb) shared by roverside

8. Run all the cells in the notebook before starting with Model building.