**Quora**   
  
**Problem** : How to recognize similar type of questions?

Ex: How ipl spoiling cricket? Or, do ipl has any bad effect on cricket?  
both are same .Can be regarded as one question  
Make the end-to-end nlp pipeline  
  
1.Data acquisition:

-quora ques ans download

-manually from quora or to ask from quora to share their data

2.Text Preparation:  
  
-cleaning : comma,semicolon other punctuations.focus on the keywords.

-remove wh word(usually there will be how/what always)

3.Feature Engineering

Positive word| negative word| (how/what/where..) | label – wh+topic+what asked  
4.Modelling  
  
Algorithm to solve: machine learning based I think.As there is not huge data.So feature has to be made manually then ml applied.  
by matching the keyword and aspect similarity,like (with cricket and aspect bad,also cricket and spoil (negative aspect),so both are same)  
evaluation-

Intrinsic-normal testing

Extrinsic-user asking a ques,whether my model showing similar question or not  
  
5.Deployment  
-in web or quora app

-monitor the result whether evaluation result is satisfied

-update needed then change it