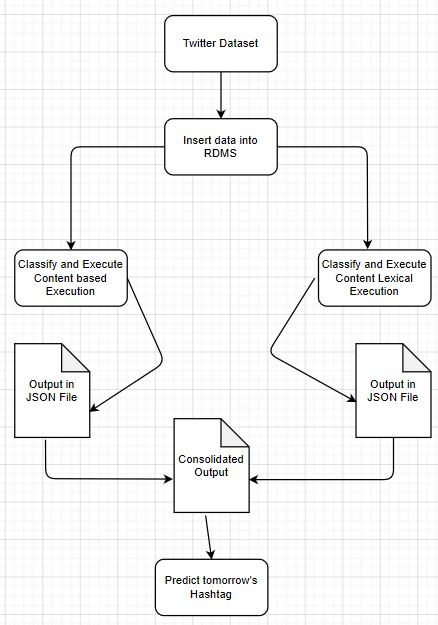
Here’s the Work Flow,



Explanation in Points,

1. From the downloaded datasets, it’s clear that we have three components in it – T, U, W
   1. T – Timestamp
   2. U – Twitter User URL
   3. W – Tweet which user has done.
2. We classify the components which are available above and derive the information into Relational Database – for now using MySQL DB.
   1. Hashtag – hashtag that were available in the tweet.
   2. Mention - user that were mentioned in the tweet.
   3. Id – primary key of the record – can also be considered as tweet\_id
3. Execute the Content python to get output in JSON.
4. Execute the Content Lexical python file to get output in JSON.
5. Executing final file to compare results that occurred between hashtags and score them accordingly.
6. Predict them and show it in graph.

Questions from Professor,

*My professor asked me two questions: How would you evaluate the performance of algorithm. In other words: what does it really mean that the hashtag is popular or not - would it based on some ground truth and if so, on which one?*

1. Time taken to execute Content Python File is 2-3Minutes in total.
2. Time taken to execute Content Lexical Python File is 5-8 minutes because, there are n number of queries that needs to be executed to find AveAuthority
3. So overall, Database size for holding 1M records is 187 Mega bytes
4. Therefore, getting data from DB by running queries is happening between 100ms to 1 minutes – which depends on data weight in retrieval rate.