# WEB SCIENCE
# COURSE WORK 1
# Matriculation Number -2576183S

1. **How would you identify and crawl public reaction to BBC news or other emerging events from tweets? How would you organise these for the use of BBC? [Hint: the problem is to identify tweets related to news and public reaction to it – we need to identify them and organise them for public consumption]**

Since, we need to extract public reactions to BBC Scotland news, we can capture public thoughts and opinions by identifying –

1. The Quoted retweets of BBC news's tweets.
2. The Replies to the BBC news's tweet.
3. Tweets containing hashtags and important query terms related to the trending news.

Following steps to Crawl the Twitter -

Using api.user_timeline first extract all the latest tweets (let's say 100) by the BBC Scotland site, screen name or user id of BBC Scotland is passed as a parameter to user_timeline api.

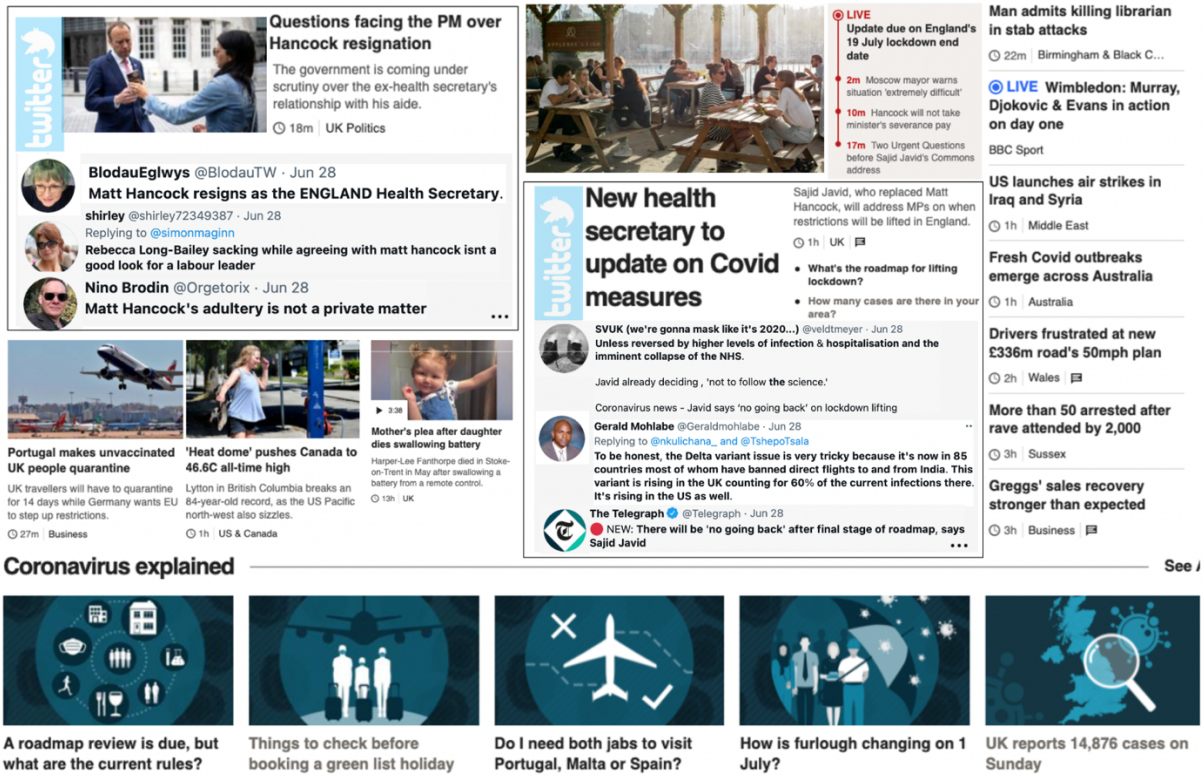
Now, from the tweets returned

- corresponding quoted retweets are extracted using api.retweet
- replies to the tweets are extracted
- query terms and hashtags from entities object are extracted, stop words are removed from this set of query terms, and after prioritizing highly bursting terms the related tweets are searched with api. Search. Therefore, the most recent tweets that were relevant to the query terms are found.

Also, tweets and replies from priority users are given higher rank, identification of priority users is done by using taking average of account's age weight, followers' count and verified status weight.
After all the tweets are extracted, duplicates, noise and information less tweets are removed.
Out of all the tweets extracted, tweets related to same news are grouped using content analysis by clustering similar content using indexing.

Using a priority ranker with a frontier queue, the scheduler is prioritised and process of collecting data is optimized.

2. **DESIGN a web site to incorporate public feedback? (Please provide a design sketch along with your justification)**




Displaying public feedback only for trending news or you can say news being discussed. Only displaying three top tweets of a news, a user can see more related tweets by clicking on the ellipses in the lower right corner.

By clicking on any tweet, a user will be directed to the tweet, so that he can see more on the discussion.

This page is more like a bulletin page, where important news along with discussion going on it is displayed. A user cannot add its own comment here or like/dislike any news.

3. **How would you design an A/B testing procedure for incremental development? [Hint – identify hypotheses, design and measures? Discuss internal and external validity issues]**

   We test how does the feature, "including public feedback in terms of tweets" affect the web Page performance. So, this is the feature to be tested.

The goal is to identify that if including public feedback improves the user experience in terms of following metrics -
User engagement–

- Lower bounce rate.
- If user clicks the eclipses to read more public feedback i.e., increase click count

User loyalty –

- no. of user sessions i.e., how often a user visits the page.

Hypothesis –

H0: Including public feedback will not impact the bounce rate, click count and user sessions.

H1: Including public feedback will decrease bounce rate, increase click count and increase user sessions.

We create two pieces for the BBC Scotland page. The existing design without public feedback is "Control" (version A) and the design version in question 2 is the variation on version A, it is a "Treatment" (version B).

Both these two versions are exposed to two classes of potential online site users, 50% of users will see version A and rest 50% will see version B. The site visitors are split randomly, and the feedback from both control and treatment users is recorded, based on which we can decide which version performs better.

Both variations are run and tested simultaneously.

3 weeks will be the ideal time to run the test because both user engagement and loyalty are being tested and user loyalty reacts slowly. This period is long enough to make accurate conclusions about the results. Temporal effects like holidays or if there is a big event occurring like elections should be considered.

To measure results – If rate of success differs across two groups, then Chi-squared test is used to determine the statistically significant difference between expected frequencies and obtained values. Also, need to check for statistical significance for any sampling error. Statistical significance indicates, to a degree of confidence, the likelihood your test findings are reliable and not due to chance.

Validity has to do with whether other factors outside of sample size are affecting your data negatively.

Internal Validity –

statistical violations cannot be sure of the AB test results only by reaching statistical significance. Also need to reach the required sample size as well. Otherwise, the results might just be imaginary. Residual effect because Using a user split exactly same as past experiments. Participation in the pre-test influences the reaction to the 'treatment'. Participants in the control and experimental group differ substantially and can thus not be compared.

External Validity

The new features implemented later might affect the results obtained in this test. The results shown in experiment still hold when new features are rolled out to entire population. Experimenting on a sub population and assuming impact on general population. The bias or error that occurs because the sample is not representative of your entire audience. Time effect is also one of the threats to external validity.

4. **If you were to use crowdsourcing-based evaluation, how would you do it? What benefits or drawbacks you see in using crowdsourcing-based evaluation? [Hint - Design a crowdsourcing-based study; what questions to ask? In what order and what benefits and drawbacks compared to A/B Testing?]**

   Identify the goal or Problem –

   The goal is to see if inclusion of tweets to the news enhances user's experience. The problem can be divided into small and simple tasks and multiple workers can work independently. The workers can be paid. The tasks are defined such that the results of the workers are measurable.

   Plan the content for the micro tasks –

   This task is partitioned in smaller tasks called as micro tasks to reduce the amount of time, uncertainty, the cost and the procedure and the results to be more effective. Then all the answers are collected.

   First, problem is decomposed into smaller tasks operating simultaneously by many workers

   The purpose is that a user should get to know if a tweet is related and relevant to the news or not. We show multiple news with multiple tweets and ask user which is more relevant to the tweet and which is not. So here workers start with searching the relevant tweets and then make comparisons about the relevance between the initial website and the new one, which provides tweets.

   Later we can ask questions related to design aspects. If location of tweets are significant, or colour and including a twitter logo helps or not.

Bias can bring only drawbacks, so to avoid biased answers we can just hide the topic that is represented each time, and the workers answer the questions without knowing the topic.

Age, gender, location could be considered as bias, so we can gather this information and show certain groups certain types of news provide the results based on this.

Determining the sub-tasks –

The goal is to identify which tweets are relevant and irrelevant to the news and to provide some answers about the relevance (interface task). The questions asked to the users are the following:

The options available for all questions is i.e., a user can select one of the following five options

i) Strongly agree
ii) Agree
iii) Neutral
iv) Disagree
v) Strongly disagree

- Do you think that three tweets are enough for each news?
- Do you think that the tweets are relevant to the news?
- Do you think location of the tweets with respect to the news is helpful?
- How often do you click the ellipses to see more tweets?
- Do you think that all the news on the front page should include their own tweets?
- Do you think that including twitter logo grabs the user attention?

Select the crowdsourcing platform to use

The task is made available to the worker using Figure eight crowd sourcing platform

publish

We publish the task and wait for the workers to respond. However, workers should be evaluated to reinforce the correctness of the results. We need to analyse accuracy of results obtained, this can be done by looking at threshold values and comparing.

Benefits and Drawbacks compared to A/B testing:

- A/B testing reacts slowly; faster results could be obtained from crowdsourcing.
- A/B Testing is done on real users, but crowdsourcing is done on paid/unpaid users, but quality of the results could be affected in crowdsourcing as the users could be spammers and due to the other personal issues of the user might affect the result.
- Better quality of results is obtained from A/B testing as one can control the sample and users in the sample because human workers are slow and more expensive and can give incorrect answers.
- Crowdsourcing could be more costly when compared to A/B testing for example in the above test scenario as there is no cap on the number of users could attempt the task this could increase the budget.
- Humans might get confused while running complex work/tasks.