* **What is the problem you want to solve?**

The problem is based on Text Classification of the News articles. The goal is to categorize new articles into multiple classes as per its type, given heading and short description about news articles. .

* **Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis?**

Client is Huffpost, a US based company which reports news through online medium, on their website. One of main functionality of Huffpost is, they take inputs from readers and covers only those topics readers are interested in. Therefore, those news needs to be categorized as per its type.

It would be best for the company if they could build an algorithm which could categorize news article on the basis of only news headline and short descriptions about the headline. To build an optimized algorithm for this problem, Huffpost only have to input very basic details about the news such as, news heading, short description of the news, etc. It will be categorized automatically without any involvement of human to read news and categorizing it manually. This will boost the process of reporting fresh news on portals.

* **Briefly outline how you’ll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution**

Problem identification: Prepare a problem identification report.

Data wrangling: Loading data, structure it by cleaning data.

Exploratory data analysis: An EDA to identify list of words which appeared most in particular categories. Plotting basic bar graphs and plot to identify count of words with respect to its category.

Pre-processing and training data development: Formulating hypotheses regarding individual factors (features) for conducting correct data clearing and data preparation for modeling.

Modeling: Building a pipeline for text classification process. Model the data with available features and then tune the model.

Documentation: Conclusions and recommendations.

* What are your deliverables?

A GitHub repo containing the work completed for each step of the project, including:

■ A slide decks

■ Project reports

* What data are you using? How will you acquire the data?

Using basic details of already published news articles data such as, news heading, short description of the news, etc. to build an algorithm. Data is sourced from online portal Kaggle.