**Problem Statement 17**

* Due Jul 9 by 23:59
* Points 10
* Submitting a file upload
* Available Jun 18 at 0:00 - Jul 9 at 23:59

**General Instructions:**

1. Follow the instructions in each question carefully.
2. Each group is expected to submit jupyter notebook (.ipynb) with output for each cell.
3. Submissions using other python IDEs will not be considered for grading.
4. In case the link to dataset is not useful, same dataset can be downloaded from any online resource.
5. A clear explanation for each output obtained is mandatory.
6. Contribution of each member of the group must be mentioned.
7. Justification of the output obtained for all the tasks is mandatory.

**Link to the Dataset:**

The following data is intended for advancing financial sentiment analysis research

<https://drive.google.com/file/d/15Wmi3GyCeEX6w8N528S40OYCnRhInH5W/view?usp=sharing>

[Links to an external site.](https://drive.google.com/file/d/15Wmi3GyCeEX6w8N528S40OYCnRhInH5W/view?usp=sharing)

**Description of Data:**

The data set is intended for advancing financial sentiment analysis research

1. Download the dataset and Create a dataframe named as **fsa,**then check the head, info, and describe methods on created dataframe **fsa**.**(1 Mark)**
2. Perform pre-processing steps like Removing Punctuations, Numbers, and Special Characters, Stop Words in dataset.  **(1 Mark)**
3. Normalize Sentence column by using Stemming or Lemmatization. **(1 Mark)**
4. Preprocessed sentence should be included in the **fsa** data frame as ‘cleaned\_sentence’. Plot word cloud for the sentence. **(1 Mark)**
5. Create two objects **X**and **Y**. **X** will be the ‘cleaned\_sentence' column of **fsa** data frame and **Y**will be the ' Sentiment' column.**( 5 Marks)**
   1. Create a Skip-gram object and split the data into training and testing sets. Train a Decision tree model. Display the confusion Matrix for both train and test.
   2. Create a CBoW object and split the data into training and testing sets. Train a Decision tree model. Display the confusion Matrix for both train and test.
   3. Compare Skip-gram and CBow. ***Answer without justification will not be awarded marks.***

6. Display the HMM POS tagging on the first row of ‘cleaned\_Review’. **( 1 Mark)**