REPORT on obstacles:

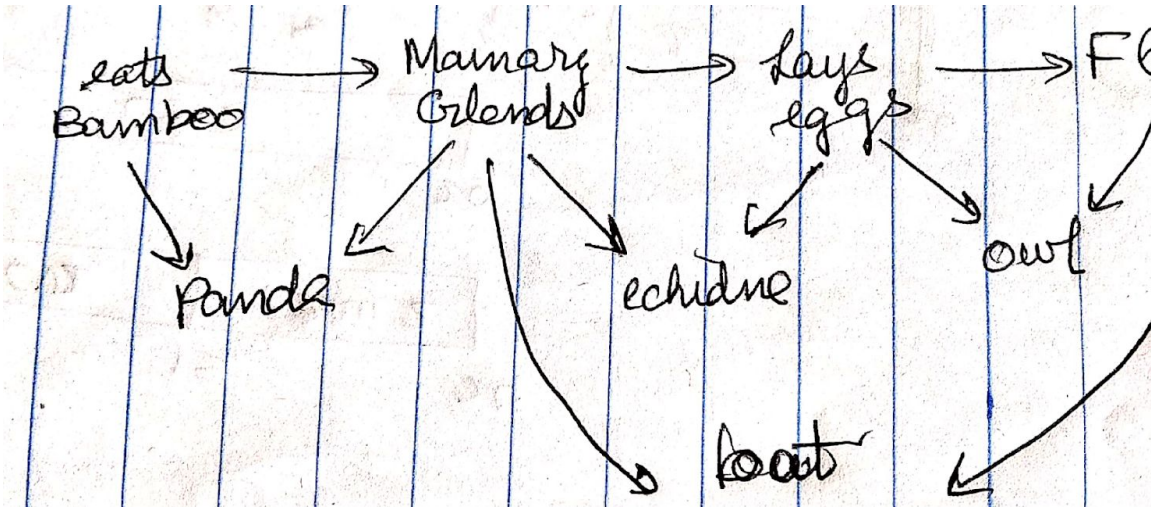
1. Selection of models for training:
    - Selection of naive bayesian:
        - Difficulty in converting the given data into the form of matrix required by naive bayesian
        - Used different text-mining methods, nltk to scipy but faced different problems with each
        - Import text-mining method
            - was removed in python 2.
        - On returning a matrix:
            - <textmining.TermDocumentMatrix object at 0x00000250F1F42C88>
            - Gets returned as object rather than matrix
2. While preprocessing
    - have to split the first word and the rest of the words, write , when done as bytes:
    - needed words to be converted from string to bytes(solution)
3. Implementation platform:
    - Implementation in collab required the dataset to be in the google drive and navigation to a specified zip folder and authorization on every insert needed to be approved by the drive user.
    - Python in pycharm had an issue with text-mining packages.
    - Hence chose notebook(solution)

4. Final implementation problems:

    - For any new word passed to the model trained with Word2Vec, showed words not present in vector space error.
    - This is expected since the word2vec model takes all the words to generate a model and scores each word based on the proximity to the axes in the vector space
    - Solution:
        - While building passed all the words seen so far as we see more data in the training phase . Training the model happens purely with the training data passed in that iteration . Hence when passing the test data, the model was able to return predicted vectors based on the model built with train data.

5. Learnt about classification problem and basic terminologies, was unclear about the multi-class and multi-label classifier:
    - Feature: individual measurable property

        - Binary and multilabel classifier: number of classes in the expected output
        - Binary and multiclass classifier: 2 or more than 2 classes
    - Unfamiliarity with therms multivariate, domain-theory:

- In Question 1,
    - Was unsure about the ordering of the nodes!
    - When tried with causal effect!
    - Got the following graph:
    - 

- A very complex network