Sample questions for CIE2 (Not limited to these)

Unit 2:

[6. Learning to Classify Text (nltk.org)](https://www.nltk.org/book/ch06.html)

1. Explain supervised classification with a neat diagram with gender identification
2. Demonstrate with code, the classification of names as male or female by training the names corpus (refer textbook and textclassification.ipynb file in drive)
3. Explain overfitting with example
4. Explain document classification. Write python code to demonstrate classification of movie reviews as positive or negative
5. Write python code to demonstrate POS tagging using python code using brown corpus. Show accuracy, train set and test set
6. Write python code to demonstrate sentence segmentation (sentence boundary detection)
7. Explain importance of accuracy, precision and recall, confusion matrix and cross validation
8. Explain the working of decision tree classifier with a note on entropy and information gain
9. Explain the working of Naïve Bayes algorithm

Unit 3

[7. Extracting Information from Text (nltk.org)](https://www.nltk.org/book/ch07.html)

1. With a neat diagram explain the information extraction model
2. Given a document, demonstrate preprocessing of the document with python code (refer .pynb file)
3. Differentiate between chunking and chinking with example
4. Write python code to demonstrate regular expression based NP chunker. The given sentence should be preprocessed and parsed as per grammer. The output should be a tree
5. Explain Named Entity Recognition with an example application
6. Explain term frequency and inverse document frequency with example

[8. Analyzing Sentence Structure (nltk.org)](https://www.nltk.org/book/ch08.html)

1. Explain CFG and write a simple grammer to parse the sentence “the cat drank the milk”
2. Define recursion in syntactic structure
3. Using recursive descent parsing parse the sentence “the dog saw a man in the park”. Show steps
4. Using shift reduce parser, parse the sentence “the dog saw a man in the park”. Show steps
5. Write python code to demonstrate CFG. It should take sentence as input and draw tree
6. Write python code to build a spellchecker using your own dictionary