**ASSIGNMENT: NLP based problem to estimate difficulty level of kids stories**

We are sharing text data which we have extracted from stories used in the Freadom app. The stories in the Freadom app need to be simple and easy to understand. To ensure that student gain most from each reading session, it is imperative to make sure that the stories which we offer to student is at his level of proficiency. If there is conflict between the level of proficiency of student with that of the story which we are recommending to students, it will not lead to any learning or improvement in reading proficiency.

To determine if the story is at the appropriate level, multiple factors are taken into account and one of them is the structure of the story text itself which depends on-

* Overall length of the story
* Average length of the sentences in the story
* Whether the words used in the story are familiar or not
* Whether the words used are simple or complex etc.

Carry out text processing on the attached .txt files to develop a model which estimates the relative difficulty level of these stories basis their text data only. The code should take separate text files as input and give out a final table where each row represents one text file and contains all the features you have developed along with final difficulty score. Please also explain what each feature means and how it has been calculated either within the code or in a separate word doc.

You are free to use any approach / model as you see fit. Your assignment will be evaluated on all the steps involved in developing the model, soundness of the overall approach, any mechanism to cross validate the output etc. and not only on the final outcome.

Key points to keep in mind

* You can use any programming language or tool to develop the model but Python will be preferable
* You will need to submit the complete code including all the steps involved
* You are expected to submit a note on how you went about developing the model and any related insights