**In this code we have used Topic modelling with Latent Dirichlet Algorithm (LDA) using Gensim and pyLDAvis for visualization. The Flow of code is as follows:**

1.We have imported the various python libraries required in the first part of code.

2. Then, we have created functions for preprocessing the text such as tokenize for generating tokens,

get\_lemma for lemmatization to extract the root word and we have used nltk library for stop words removal and noise filtering.

3. Then, we have read the file and preprocessed the text from it to train the corpus.

4.we have created the dictionary and corpus and saved it in our local folder.

5. we have created our lda model using our trained corpus and dictionary.

6. In the last part of code we have loaded our model and used pyLDAvis to generate the plot for topic visualization and to predict the topic.

7. The circles area shows the dominance of the topic in our dataset and distance between circles show the inter topic distance between them also when we select a particular topic we can see which word contributes how much to that topic according to its term frequency and weight.

8. we also have lambda to adjust the relevancy (relevance metric) of the topic and to predict what the topic is about.

9.Also our model shows the top ranking words in our dataset according to their weights.

10.We can also see in the output the percentage of the token occupied by that topic. e.g : Top-30 Most Relevant Terms for Topic 3 (12.2% of tokens) which means topic 3 uses 12.2 percent of the key tokens

11. To generate our own topic we could use Guided LDA for seeding whose documentation is mentioned here: <https://medium.freecodecamp.org/how-we-changed-unsupervised-lda-to-semi-supervised-guidedlda-e36a95f3a164>