1. Text Classification of News Articles using NLP.
   Article Id – Article id unique given to the record
   Article – Text of the header and article
   Category – Category of the article (tech, business, sport, entertainment, politics)

Consider BBC News as corpus for implementing question 1

Ans:
1.Data Collection: Obtain a dataset of news articles with the required fields (Article Id, Article, Category). In this case, you can use the BBC News corpus.

2.Data Preprocessing: Prepare the data for analysis by performing various preprocessing steps such as lowercasing, tokenization, removing stop words, stemming/lemmatization, and handling special characters or noise. You may also consider removing any irrelevant information such as HTML tags or URLs.

3.Feature Extraction: Convert the preprocessed text data into numerical features that machine learning algorithms can understand. Some commonly used techniques for feature extraction in NLP include bag-of-words (BoW), TF-IDF (Term Frequency-Inverse Document Frequency), and word embeddings like Word2Vec or GloVe.

4.Training-Testing Split: Split the dataset into training and testing sets. The training set will be used to train the classification model, while the testing set will be used to evaluate its performance.

5.Model Selection and Training: Choose a suitable machine learning algorithm for text classification, such as Naive Bayes, Support Vector Machines (SVM), or deep learning models like Recurrent Neural Networks (RNNs) or Transformers. Train the selected model using the training data.

6.Model Evaluation: Evaluate the trained model using the testing set. Common evaluation metrics for text classification include accuracy, precision, recall, and F1-score. These metrics will give you an understanding of how well your model is performing.

7.Hyperparameter Tuning: Fine-tune the model by adjusting its hyperparameters to improve its performance. This can be done using techniques like grid search or random search.

8. Prediction: Once the model is trained and optimized, you can use it to predict the category of new, unseen news articles.