

FAKE NEWS DETECTION USING ML

PROBLEM STATEMENT:

Continue building the fake news detection model by applying NLP techniques and training a classification model. Text Preprocessing and Feature Extraction, Model training and evaluation.

TEXT PROCESSING:

This step involves tokenization, removing stop words, handling punctuation, and stemming or lemmatization. We can use libraries such as NLTK or spaCy for these tasks.

FEATURE EXTRACTION:

We can use TF-IDF, Word2Vec, GloVe, or BERT embeddings to represent the text data in a numerical format that can be utilized by the machine learning model.

MODEL TRAINING:

Popular classification models for NLP tasks include SVM, Naive Bayes, Random Forest, Gradient Boosting, or deep learning models like RNN, LSTM, or Transformer-based models such as BERT, GPT, or RoBERTa.

EVALUATION:

To assess the model's performance, we can use metrics such as accuracy, precision, recall, and F1-score. Cross-validation or a holdout validation set can be used to evaluate the model's generalizability.

HYPERPARAMETER TUNING:

We can fine-tune the model by adjusting hyperparameters and employing techniques like grid search or random search to find the optimal set of

parameters that yield the best performance. The integration of a robust pipeline, thoughtful feature extraction, and model selection plays a crucial role in building an effective fake news detection system.