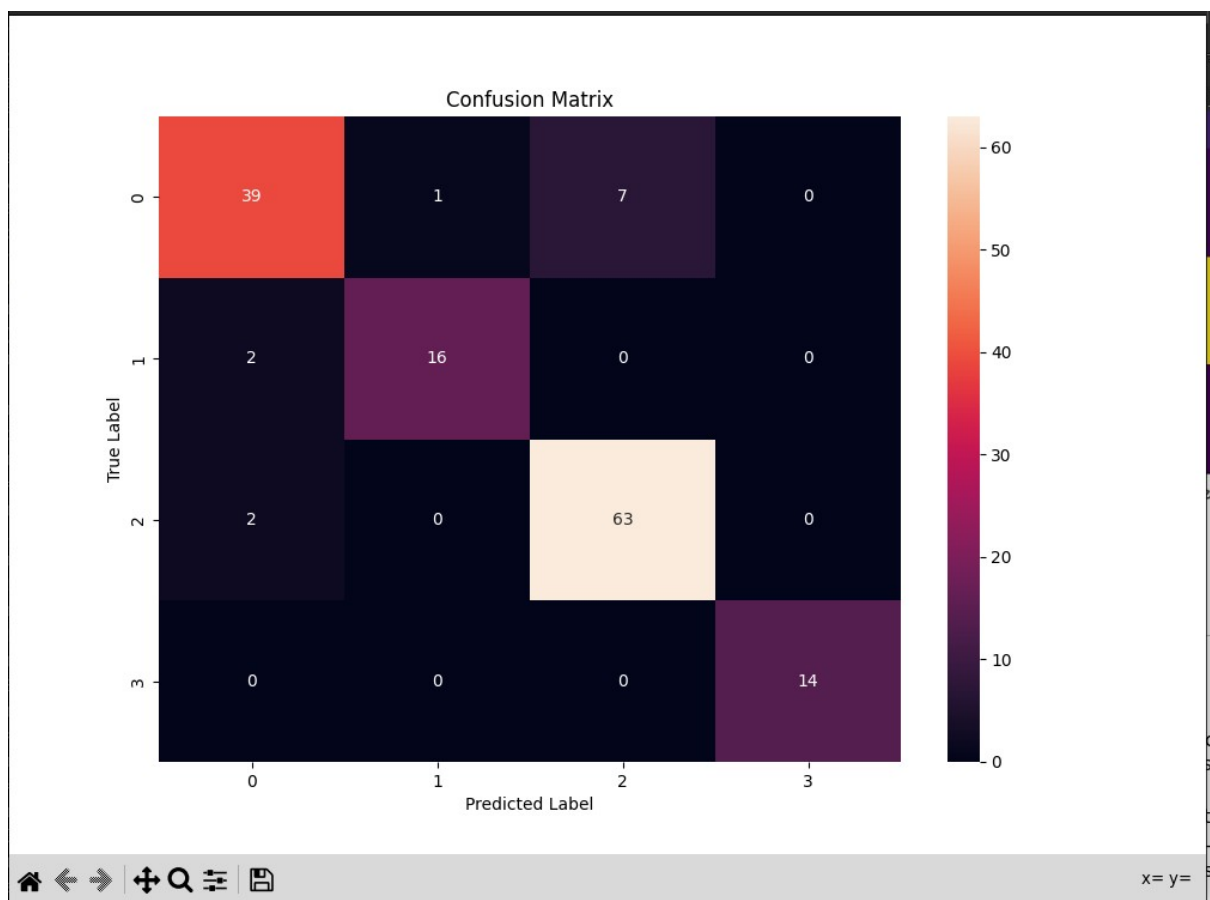


Name: Usman Akinyemi

Course: MLPR

Lab: Lab9



1. Stemming to reduce words to their root form by removing suffixes, while lemmatization involves reducing words to a common base or root.

2. TF-IDF (Term Frequency-Inverse Document Frequency): TF-IDF is a statistical measure used to evaluate the importance of a word in a document relative to a collection of documents (corpus).

3. Alternatives to TF-IDF: Some alternatives to TF-IDF include:

Word Embeddings

Bag of Words (BoW)

N-grams.

4. Advantages of word2vec over TF-IDF:

Semantic Meaning: Word2vec captures semantic meanings of words, allowing for better understanding of word relationships.

Contextual Information: Word2vec considers the context of words, providing more contextually relevant representations.

Dimensionality Reduction: Word2vec generates dense, low-dimensional representations, which can be more efficient than the sparse representations of TF-IDF.

5. Applications of the Lab Assignment Problem:

Text Classification.

Information Retrieval