**Assignment Problem:**

The assignment involves leveraging the Google Speech Command Dataset, a subset of which comprises 105,829 one-second audio files with utterances of four common words ("Yes," "No," "Stop," and "Go"). The goal is to perform speech recognition using machine learning techniques.

**Solution:**

(a) Data Preparation and Convolutional Neural Network (CNN) Training:

Loaded audio spectrograms and class labels from XSound.npy and YSound.npy, respectively.

Split the data into training, validation, and test sets.

Reshaped data to fit the CNN model.

Trained a CNN with hyperparameter tuning for optimal performance.

(b) CNN Performance Analysis:

Achieved a test accuracy of approximately 94.6%.

Utilized a confusion matrix and classification report for detailed model evaluation.

(c) Long Short-Term Memory (LSTM) Model Training:

Implemented an LSTM model, known for capturing temporal dependencies in sequential data.

Trained the LSTM model and evaluated its performance on the test set.

(d) LSTM Performance Analysis:

Obtained a test accuracy of around 95.2%.

Visualized the confusion matrix and generated a classification report for comprehensive evaluation.

**Reflection on Learning Outcome:**

Both CNNs and LSTMs were effective in audio classification, with CNNs capturing spatial patterns and LSTMs excelling in modeling temporal dependencies.

Considerations for model selection involve the nature of the audio data, the need for capturing long-term dependencies, and the dataset size.

LSTMs proved valuable for sequential audio data, while CNNs excelled in capturing local patterns in spectrograms.

The assignment provided practical insights into choosing appropriate models for audio classification tasks and understanding their strengths and limitations.