

Tanya Chopra  
102116069  
4CS11  
Email: tchopra1\_be21@thapar.edu  
SESS\_LE1

# Report

**Paper summary** :The paper introduces the Speech Commands dataset, designed to advance keyword spotting systems by providing a standardized dataset of spoken words. This dataset allows researchers to build and test small models for detecting specific words amidst background noise. Released under the Creative Commons BY 4.0 license, it aims to support reproducible research and comparisons in keyword spotting, addressing the need for accessible and specialized datasets beyond traditional speech recognition resources.

**Statistical Analysis**:The dataset used for training, validation, and testing has the following characteristics:

- **Training Set Lengths**: The average length of audio samples is approximately 15,694 samples, with a standard deviation of 1,141 samples, indicating some variability in the sample lengths.
- **Validation Set Lengths**: The average length is slightly less at about 15,693 samples, with a standard deviation of 1,154 samples, showing similar variability to the training set.
- **Test Set Lengths**: All audio samples in the test set are uniformly 16,000 samples long, with no variability, suggesting consistent audio lengths for testing purposes.

These statistics reveal that while training and validation sets have a range of audio sample lengths, the test set is standardized to a fixed length.

## Model Performance Summary:

### Model Architecture:

- **LSTM Layers**:
  - **LSTM (64 units)**: Outputs a sequence of shape (None, 16000, 64) with 16,896 parameters.
  - **LSTM (64 units)**: Outputs a vector of shape (None, 64) with 33,024 parameters.
- **Dense Layers**:
  - **Dense (128 units)**: Outputs a vector of shape (None, 128) with 8,320 parameters.
  - **Dropout**: Applied with no additional parameters.
  - **Dense (12 units)**: Outputs a vector of shape (None, 12) with 1,548 parameters.
- **Total Parameters**: 59,788 (233.55 KB), all trainable.

### Training Performance (Epochs 1-2):

- **Epoch 1:**
  - **Training Accuracy:** 63.13%
  - **Training Loss:** 1.5842
  - **Validation Accuracy:** 62.15%
  - **Validation Loss:** 1.5634
- **Epoch 2:**
  - **Training Accuracy:** 63.42%
  - **Training Loss:** 1.5354

The model shows an improvement in training accuracy and a slight reduction in training loss from Epoch 1 to Epoch 2. The validation accuracy and loss provide an initial benchmark for performance, indicating a relatively close fit between training and validation data and the project continues further in progress for evaluating and fine-tuning the model.

**New Dataset link:** [https://drive.google.com/drive/folders/1R4W0HdWjg3K2UETLC9-xt-9RErB7\\_V\\_q?usp=sharing](https://drive.google.com/drive/folders/1R4W0HdWjg3K2UETLC9-xt-9RErB7_V_q?usp=sharing)..created as of now ...

*\*\*And .ipynb file is part of the repository*