Karibuni

Swahili Audio Prediction

Meet the team

Manyara Baraka
Nangi Mugira
Collins Kanyiri
Benson Kinyua
Jacinta Mukii
Faith Nyawira

TABLE OF CONTENTS

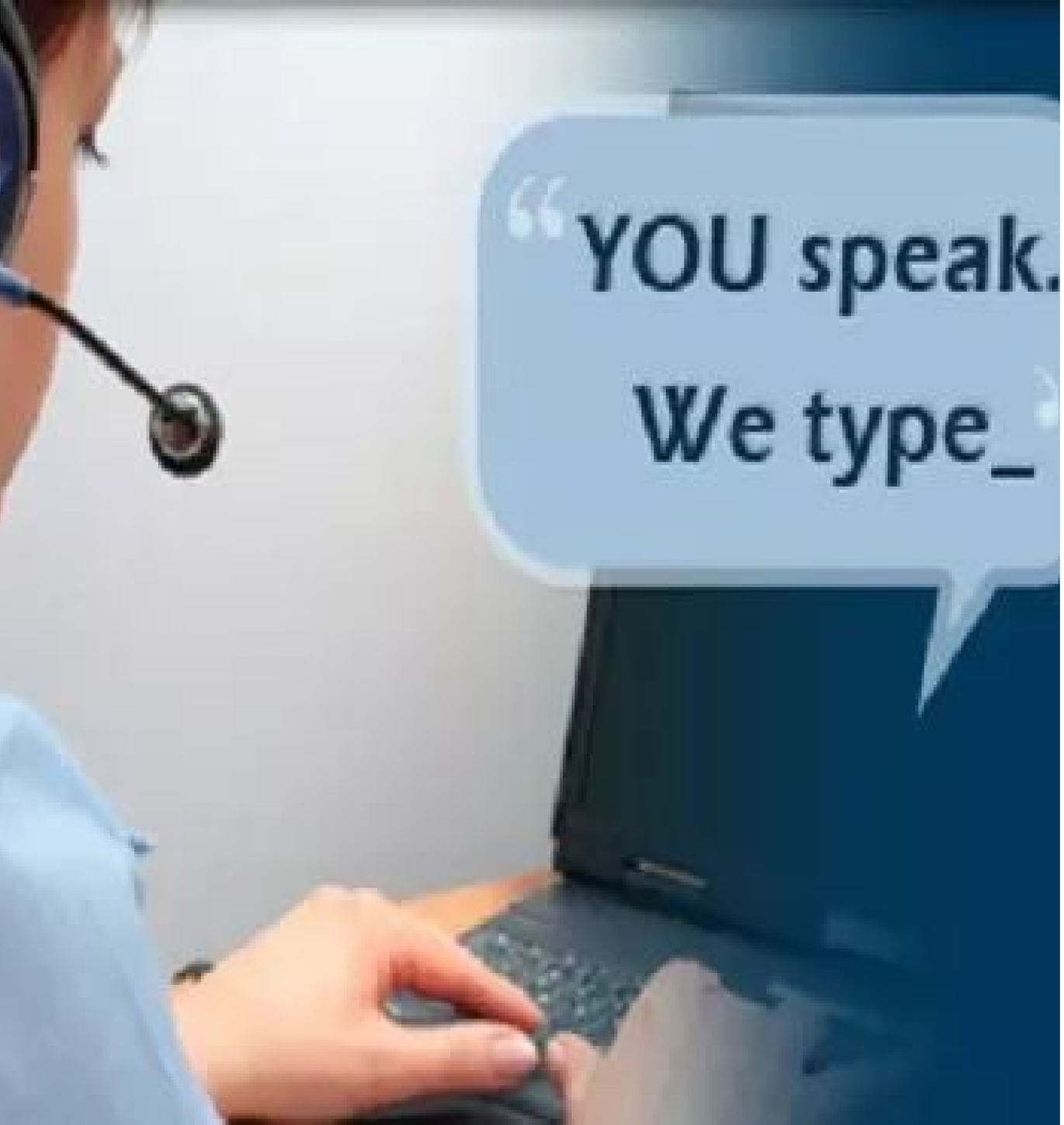
Why is Swahili important to us?

- Facilitates communication in many African regions.
- Fosters cultural connection among African communities, transcending borders and differences.
- Preserves the Historical and Artistic Heritage, i.e, storytelling, music, and various artistic expressions.

Problem Statement

- We aim to empower the Swahili language through technology to ensure its growth, cultural preservation, and enhanced accessibility in the digital age.
- As the accessibility of digital audio content in Swahili continues to grow, the development of an automated transcription system for Swahili audio recordings becomes essential.
- This initiative will yield benefits for a range of stakeholders:

Content providers, broadcasters, and telecommunication companies.
Language learning platforms.
Researchers in linguistics and African languages.
Swahili-speaking communities and individuals.



Main Objective

To develop an automated system for converting basic Swahili audio into written text using speech recognition technology.

Specific Objectives

01

To develop a machine learning model capable of transcribing Swahili audio recordings.

02

To deploy a model that transcribes the recorded audio files.

03

To provide recommendations for further enhancements and applications.

Data understanding

The data used in this project was collected by 300 contributors based in Kenya. It consists of recordings of twelve different phrases spoken in Swahili. Here are the 12 words and their English translations. We are predicting the Swahili words; the English translations are here for interest's sake.

Swahili	English
ndio	yes
hapana	no
moja	one
mbili	two
tatu	three
nne	four
tano	five
sita	six
saba	seven
nane	eight
tisa	nine
kumi	ten

Data modeling

We have used two models for this project:

- The **AlexNet** model exhibited a notably low performance, achieving only a **6% accuracy** rate.

- **Resnet18 Model** which had the best performance with **WER = 6%** and **accuracy of 94%**.

Model interpretation

- A 95% accuracy rate has been achieved by the Resnet-18 model, indicating that 95% of the tested words have been correctly predicted.

- A "WER" (Word Error Rate) of 5% achieved by the Resnet-18 model means that 5% of the words in the transcribed text are incorrect or contain errors.

Conclusions

ResNet18 ASR Model

Training Loss: The training loss for the ResNet18 ASR model decreased over the epochs, indicating that the model learned the data well. Like the AlexNet model, the validation loss is consistently higher, indicating a potential overfitting issue.

Accuracy: The accuracy for the ResNet18 ASR model is much higher, around 0.937, which is a significant improvement compared to the AlexNet model. This suggests that the ResNet18 architecture might be better suited for the ASR task.

Word Error Rate (WER): A WER of 0.063 indicates that the ResNet18 model is making relatively few errors in transcribing the audio recordings. This is a positive sign of its performance.

Recommendations

1. Further Model Evaluation: While ResNet18 shows promising results, it's essential to perform a more rigorous evaluation, including testing on a larger and diverse dataset.
2. Incorporate a broader range of Swahili audio recordings, including longer sentences and passages to expand the scope of this project.
3. Language Model Integration: To enhance transcription accuracy, consider incorporating language models to improve the fluency and coherence of the transcribed text.
4. Data Quality and Preprocessing: Ensure that the data used for training and validation is of high quality and that proper preprocessing techniques are applied. Data augmentation methods can also be employed to enhance the model's ability to handle various audio conditions and audio formats.
5. Use a powerful GPU device to train the model on big audio files and improve the model's performance.

Asante.

Jisikie huru kuuliza maswali yoyote.