

## Introduction

This report summarises the development and evaluation of our Automatic Speech recognition model trained on a Twi dataset with a focus on financial vocabulary. To achieve this ASR model we utilised the pretrained OpenAI's whisper model on a custom dataset , preprocessing the data and evaluating the performance. The model is deployed via Hugging Face API.

### Work Process:

#### 1. Dataset Preparation:

A custom twi dataset which was given to us containing the training and testing datasets , accompanied with the respective audio files. The training and testing datasets were preprocessed by encoding the .ogg audio files.

But first we prepared our data\_set by

- Filtering out the rows with bad files
  - Include image
- Normalizing text to upcase to standardize the input and output
- Shuffling and splitting the Train dataset into training which was 80% of the dataset and validation which was 20% of the the training dataset.
- Use then used the Whisper process for the feature extraction and tokenizer

#### 2. Model Fine-Tuning:

The whisper small model was fine tuned on the training dataset. We used the whisper small because it balances between the accuracy and efficiency trade-offs. The whisper small has better transcription accuracy compared to the base and tiny models and provides a better accuracy without excessive computational demands unlike the medium and large models.

For training the below were used:

- A learning rate of  $1e-5$
- And the trained model would be saved at steps of 1000
- Warmup steps of 50 to stabilize the initial learning
- A batch size of 16 for GPU optimization
- The training epochs of 20
- We also included metrics of word error rate(WER) and character error rate(CER) to evaluate the character level transcription accuracy.
- 
- Save steps at 200 , the save steps are the checkpoints in which the model is saved during training so that if any interruptions occurred

#### 3. Evaluation

Whilst training after each 200 save steps the model would be evaluated with the validation set which was 20% of the the training dataset. From there the model is tested using the 10% dataset. We then finally used the 20 examples that we made to evaluate the model.

For our first (checkpoint -1400) model:

The below was our test results:

```
File 1/20: twi/New Recording 104.m4a
Predicted: Mepaa'kyew mekanea'kyew cedis
Reference: Mepa wo kyew can i get cedi mmienu
```

```
File 2/20: twi/New Recording 105.m4a
Predicted: Me sika no asa
Reference: Sika no asa
```

```
File 3/20: twi/New Recording 106.m4a
Predicted: Me pese me tɔ gu
Reference: Mepe se me kɔ bank
```

```
File 4/20: twi/New Recording 107.m4a
Predicted: Nnipa bebre pa a agyina
Reference: Nnipa bebre empe aduruma
```

```
File 5/20: twi/New Recording 108.m4a
Predicted: Kamene wo ye nkɔmmɔ
Reference: Come me ne wa yɛn kɔ bank
```

```
File 6/20: twi/New Recording 109.m4a
Predicted: Mtn na mtn ho
Reference: NDC enɔ NPP, Who would win
```

```
File 7/20: twi/New Recording 110.m4a
Predicted: Maa'kɔm Enɔ ene
Reference: Baako plus mmienu is mmiensa
```

```
File 8/20: twi/New Recording 111.m4a
Predicted: Yera de
Reference: Ewurade gyae me
```

```
File 9/20: twi/New Recording 112.m4a
Predicted: Bra ha sesa
Reference: Bra ha sesea
```

```
File 10/20: twi/New Recording 113.m4a
Predicted: Me pese me tɔ credit na me de gu mtn number so
Reference: Me pe se me kɔ bank no sesea
```

```
File 11/20: twi/New Recording 114.m4a
Predicted: Mepaa'kyo didi ma me
Reference: Me pa wo kyew didi ma me
```

```
File 12/20: twi/New Recording 115.m4a
Predicted: Akwadaa me pe tɔ ma
Reference: Akosua empe aduruma
```

```
File 10/20: twi/New Recording 113.m4a
Predicted: Me pese me tɔ credit na me de gu mtn number so
Reference: Me pe se me kɔ bank no sesea
```

```
File 11/20: twi/New Recording 114.m4a
Predicted: Mepaa'kyo didi ma me
Reference: Me pa wo kyew didi ma me
```

```
File 12/20: twi/New Recording 115.m4a
Predicted: Akwadaa me pe tɔ ma
Reference: Akosua empe aduruma
```

```
File 13/20: twi/New Recording 116.m4a
Predicted: Me sika no ye buburoo
Reference: Me sika no ye bebre
```

```
File 14/20: twi/New Recording 117.m4a
Predicted: Sika ye mada wo
Reference: sika ye ma damfo
```

```
File 15/20: twi/New Recording 118.m4a
Predicted: Kɔ buu
Reference: Kɔ abɔnte and sit down
```

```
File 16/20: twi/New Recording 119.m4a
Predicted: Akose nntɔ ma
Reference: Akosua empe aduru
```

```
File 17/20: twi/New Recording 120.m4a
Predicted: Sika ye mada wo
Reference: sika ye ma damfo
```

```
File 18/20: twi/New Recording 121.m4a
Predicted: Yera de
Reference: Ewurade gyae me
```

```
File 19/20: twi/New Recording 122.m4a
Predicted: Nnipa biribiara dwoi
Reference: Nnipa bebre empe aduruma
```

```
File 20/20: twi/New Recording 123.m4a
Predicted: Mepaa'kyew kanea'kyew cedis
Reference: Me pa wo kyew can i get cedi mmienu
```

```
1.
Final Test WER: 84.848484848484
Final Test CER:56.698564593301434
```

For the first evaluation the model did not perform quite well as it had a word error rate of 85.85% and a character error rate of 56.70%. Given this we decided to train the model again by limiting it to capital letters and the twi symbols.

```
File 2918/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiMa21-JTWbePTw-Tmp005-0Mpeeb.ogg
Predicted: Wɔ ńkwankyen
Reference: Wɔ ńkwankyen
-----
```

```
File 2919/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiFm23-MKBcLgYz-Tmp005-XCK1T6.ogg
Predicted: Wɔ ńkwankyen
Reference: Wɔ ńkwankyen
-----
```

```
File 2920/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiMa28-F02TjEj7-Tmp034-qswg0R.ogg
Predicted: ste sen
Reference: ste sen
-----
```

```
File 2921/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiFm23-MKBcLgYz-Tmp034-q0Kee6.ogg
Predicted: ste sen
Reference: ste sen
```

```
File 2871/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiMa30-SKUDJnDU-Tmp068-jABcj5.ogg
Predicted: Ma me bi
Reference: Ma me bi
-----
```

```
File 2872/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiFm21-G0meq57b-Tmp066-qbJ409.ogg
Predicted: Akwadaa no to dwom
Reference: Akwadaa no to dwom
```

```
File 3187/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiMa21-JTWbePTw-Tmp005-0Mpeeb.ogg
Predicted: Sua adeɛ
Reference: Sua adeɛ
-----
```

```
Final Test WER: 8.509428598502796
Final Test CER: 8.84811591717752
```

Without cleaning and testing on test dataset (Checkpoint - 1400)

```
/usr/local/lib/python3.10/dist-packages/librosa/core/audio.py:100: DeprecationWarning:
This function is deprecated as of librosa version 0.10.0.
It will be removed in librosa version 1.0.
y, sr_native = __audioread_load(path, offset, duration, dtype, dtype_bits)
File 20/20: /content/twi/New Recording 123.m4a
Predicted: Mɛpaa'kyɛw kanea'kyɛw cedis
Reference: MEPA WO KYEW CAN I GET CEDI MMIENU
-----
Final Test WER: 91.66666666666666
Final Test CER: 58.66983372921615
```

```

/usr/local/lib/python3.10/dist-packages/librosa/core/audio.py:184: FutureWarning:
    Deprecated as of librosa version 0.10.0.
    It will be removed in librosa version 1.0.
    y, sr_native = __audioread_load(path, offset, duration, dtype)
File 17/20: /content/twi/New Recording 120.m4a
Predicted: Sika ye mada wo
Reference: SIKA YE MA DAMFO
-----
<ipython-input-7-8e759a5757cf>:91: UserWarning: PySoundFile failed. Trying audioread in
    waveform, sr = librosa.load(audio_path, sr=16000)
/usr/local/lib/python3.10/dist-packages/librosa/core/audio.py:184: FutureWarning: libro
    Deprecated as of librosa version 0.10.0.
    It will be removed in librosa version 1.0.
    y, sr_native = __audioread_load(path, offset, duration, dtype)
File 18/20: /content/twi/New Recording 121.m4a
Predicted: Yera de
Reference: EWURADE GYAE ME

```

```

/usr/local/lib/python3.10/dist-packages/librosa/core/audio.py:184: FutureWarning:
    Deprecated as of librosa version 0.10.0.
    It will be removed in librosa version 1.0.
    y, sr_native = __audioread_load(path, offset, duration, dtype)
File 12/20: /content/twi/New Recording 115.m4a
Predicted: Akwadaa me pe to ma
Reference: AKOSUA EMPE ADUMA
-----
<ipython-input-7-8e759a5757cf>:91: UserWarning: PySoundFile failed. Trying audioread in
    waveform, sr = librosa.load(audio_path, sr=16000)
/usr/local/lib/python3.10/dist-packages/librosa/core/audio.py:184: FutureWarning: libro
    Deprecated as of librosa version 0.10.0.
    It will be removed in librosa version 1.0.
    y, sr_native = __audioread_load(path, offset, duration, dtype)
File 13/20: /content/twi/New Recording 116.m4a
Predicted: Me sika no ye buburoo
Reference: ME SIKA NO YE BEBREE

```

Without cleaning on made up dataset

```

<ipython-input-8-23e02d002ab3>:91: UserWarning: PySoundFile fa
    waveform, sr = librosa.load(audio_path, sr=16000)
/usr/local/lib/python3.10/dist-packages/librosa/core/audio.py:
    Deprecated as of librosa version 0.10.0.
    It will be removed in librosa version 1.0.
    y, sr_native = __audioread_load(path, offset, duration, dtyp
File 20/20: /content/twi/New Recording 123.m4a
Predicted: MEPAKYEW KANEA YE TO CEDIS NO YE NO
Reference: MEPA WO KYEW CAN I GET CEDI MMIENU
-----
Final Test WER: 185.41666666666669
Final Test CER: 124.70308788598574

```

```

File 14/20: /content/twi/New Recording 117.m4a
Predicted: SUKA YE MADAMFO
Reference: SIKA YE MA DAMFO
-----
<ipython-input-8-23e02d002ab3>:91: UserWarning: PySoundFile failed. Trying aud
    waveform, sr = librosa.load(audio_path, sr=16000)
/usr/local/lib/python3.10/dist-packages/librosa/core/audio.py:184: FutureWarni
    Deprecated as of librosa version 0.10.0.
    It will be removed in librosa version 1.0.
    y, sr_native = __audioread_load(path, offset, duration, dtype)
File 15/20: /content/twi/New Recording 118.m4a
Predicted: KJO ABOA TWEE NEDA SEDA
Reference: KO ABONTEN AND SIT DOWN
-----
    y, sr_native = __audioread_load(path, offset, duration, dtype)
File 7/20: /content/twi/New Recording 110.m4a
Predicted: MAA KOMMO NNYE NNWO YE NNYE SEN
Reference: BAAKO PLUS MMIENU IS MMIENSA
-----
<ipython-input-8-23e02d002ab3>:91: UserWarning: PySoundFile failed. Trying aud
    waveform, sr = librosa.load(audio_path, sr=16000)
/usr/local/lib/python3.10/dist-packages/librosa/core/audio.py:184: FutureWarni
    Deprecated as of librosa version 0.10.0.
    It will be removed in librosa version 1.0.
    y, sr_native = __audioread_load(path, offset, duration, dtype)
File 8/20: /content/twi/New Recording 111.m4a
Predicted: YSEHYIA BIOM
Reference: EWURADE GYAE ME
-----
<ipython-input-8-23e02d002ab3>:91: UserWarning: PySoundFile failed. Tryin
    waveform, sr = librosa.load(audio_path, sr=16000)
/usr/local/lib/python3.10/dist-packages/librosa/core/audio.py:184: Future
    Deprecated as of librosa version 0.10.0.
    It will be removed in librosa version 1.0.
    y, sr_native = __audioread_load(path, offset, duration, dtype)
File 4/20: /content/twi/New Recording 107.m4a
Predicted: NNIPA BUBUROO WEI PE EDI KAN
Reference: NNIPA BEBREE EMPE ADUMA
-----
<ipython-input-8-23e02d002ab3>:91: UserWarning: PySoundFile failed. Tryin
    waveform, sr = librosa.load(audio_path, sr=16000)
/usr/local/lib/python3.10/dist-packages/librosa/core/audio.py:184: Future
    Deprecated as of librosa version 0.10.0.
    It will be removed in librosa version 1.0.
    y, sr_native = __audioread_load(path, offset, duration, dtype)
File 5/20: /content/twi/New Recording 108.m4a
Predicted: KOM ME NE WO YE KOM KUBAN
Reference: COME ME NE WO YEN K BANK
-----

```

With cleaning on Test dataset on my dataset

File 600/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiMa37-BTFet0MN-Tmp105-BLiKbF.ogg  
Predicted: WOKA BOROF0  
Reference: Woka borɔfo  
-----

File 601/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiFm20-CTwUaRZ4-Tmp105-AVVeBV.ogg  
Predicted: WOKA BOROF0  
Reference: Woka borɔfo  
-----

File 602/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiMa23-GAkWUqu3-Tmp105-C0GxA2.ogg  
Predicted: WOKA BOROF0  
Reference: Woka borɔfo  
-----

File 603/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiMa25-JAFefgqX-Tmp105-jZ0VSw.ogg  
Predicted: WOKA BOROF0  
Reference: Woka borɔfo  
-----

File 604/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiMa22-ENjgxffJ-Tmp105-qQRxED.ogg  
Predicted: WOKA BOROF0  
Reference: Woka borɔfo

File 584/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiFm27-AHpAijro-Tmp132-ViDhny.ogg  
Predicted: DUM KANEA NO  
Reference: Dum kanea no  
-----

File 585/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiFm23-NAHtBqEp-Tmp132-9qLn30.ogg  
Predicted: DUM KANEA NO  
Reference: Dum kanea no  
-----

File 586/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiMa19-AAkeawcj-Tmp132-iZr7ui.ogg  
Predicted: DUM KANEA NO  
Reference: Dum kanea no  
-----

File 587/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiFm34-FAccQpEp-Tmp132-evyy2c.ogg  
Predicted: DUM KANEA NO  
Reference: Dum kanea no  
-----

File 588/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiMa22-eLxRrG7p-Tmp132-ZlDBPx.ogg  
Predicted: DUM KANEA NO  
Reference: Dum kanea no

```
File 551/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiFm29-DKwnuqiy-Tmp047-yTtVLZ.ogg
Predicted: MEPAAKYEW AANE SPOKEN
Reference: Mepaa'kyew Aane (spoken)
-----
File 552/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiMa41-KSs7K7hm-Tmp047-UYmCsF.ogg
Predicted: MEPAAKYEW AANE SPOKEN
Reference: Mepaa'kyew Aane (spoken)
-----
File 553/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiFm23-PGbBevNS-Tmp047-Zm0RhT.ogg
Predicted: MEPAAKYEW AANE SPOKEN
Reference: Mepaa'kyew Aane (spoken)
-----
File 554/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiFm24-PEhXSi7L-Tmp047-lWcmTb.ogg
Predicted: MEPAAKYEW AANE SPOKEN
Reference: Mepaa'kyew Aane (spoken)
-----
File 555/3187: /content/fisd-asanti-twi-10p/audios/AsantiTwiMa25-MK0aWumC-Tmp047-uX72VM.ogg
Predicted: MEPAAKYEW AANE SPOKEN
Reference: Mepaa'kyew Aane (spoken)
```

Cleaning with 10p test document

Link to sAPI:

[Streamlit](#)

Link to git hub

[Deep-Learning-Final-Project/app.py at main · RunyararoEs/Deep-Learning-Final-Project](#)