Sentiment Classification for Multi-Language text and image pair

- Arkadip Maitra & Soumen Halder

Department of Computer Science

RKMVERI, Belur

Introduction

- With the evolution of the Internet, social media sites, in particular, have become multimodal in nature with content including text, audio, images, and videos.
- Sentiment analysis on multimodal data is therefore a very important task.

Objectives

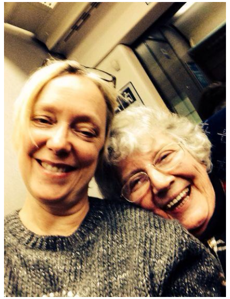
1. Creating English 100 image and text dataset from twitter.
2. Creating Bengali 100 image and text dataset from twitter.
3. Training model on publicly available MVSA dataset that contains English image and text and testing on our created dataset.

Data Processing

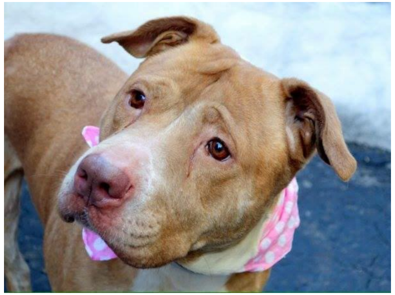
- The dataset contains image and corresponding text obtained from twitter.
- The image and text are individually labelled according to sentiment.
- The labels are , neutral(0), positive(1), negative(2).
- The labels of each image and text pair is compared to obtain the final label as follows:
  - If both labels are same then final label is the same.
  - If one is neutral, other is non neutral then final label is non neutral.
  - Else, drop data as conflicting labels.

MVSA Dataset

Some, example data from this dataset.



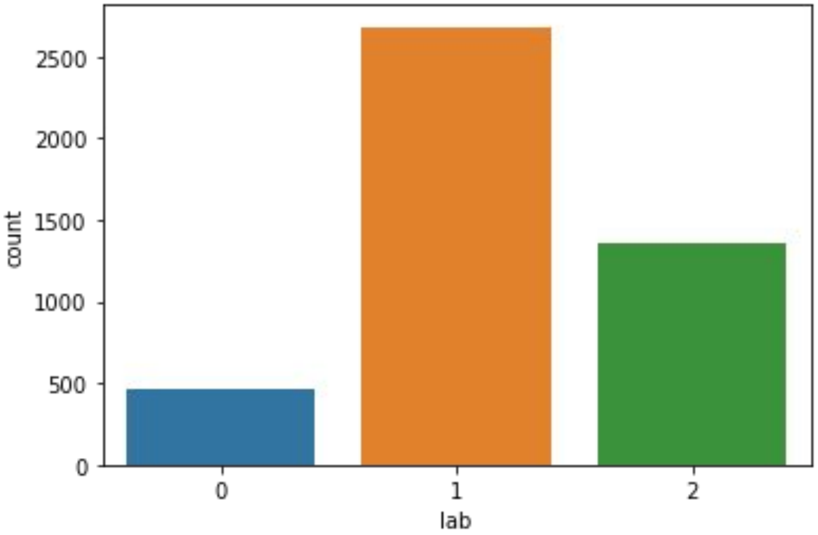
Even a packed Edinburgh to Glasgow evening train is fun in the company of ebullient art writer, Anne Ellis.



RT @southernpride50: SAFE 2/13/15. PULLED BY GLEN WILD ANIMAL RESCUE. TY. HAPPINESS HARRIET??TO DIE 2/13/15.NYC. https://t.co/5UZjZrEAZQ

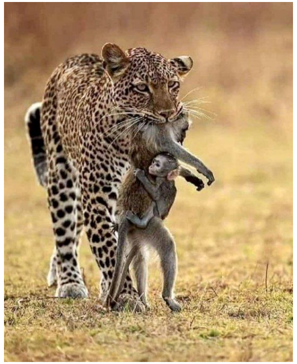
MVSA Dataset

- The distribution of the processed MVSA dataset is shown below.
- The data is unbalanced with much more positive label.

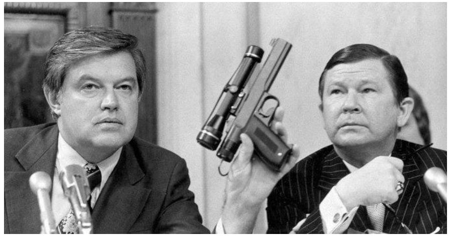


Our English Dataset

Some, example data from this dataset.



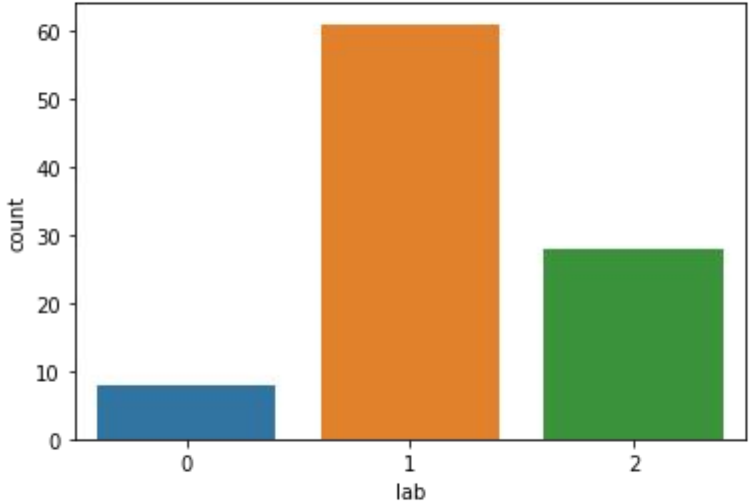
Hunger , Death and Love at the same time ..Nothing is more crueller than nature



CIA revealed a "heart attack" gun in 1975. A battery operated gun which fired a dart of frozen water & shellfish toxin. Once inside the body it would melt leaving only a small red mark on the victim where it entered. The official cause of death always heart attack @JFKSaid

Our English Dataset

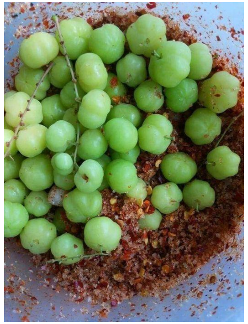
- The distribution of the processed English dataset is shown below.
- This data is made keeping in mind the distribution of processed MVSA dataset.



Our Bengali Dataset

Some, example data from this dataset.

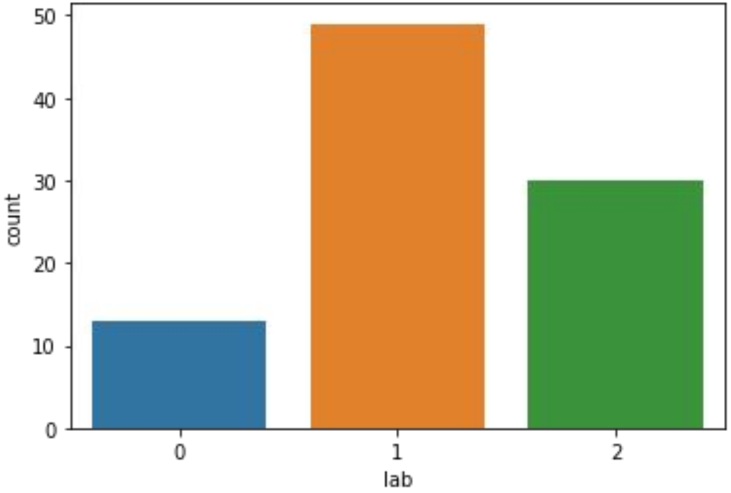
একুশ বছর আগে এই মঞ্চের ঠিক এই জায়গায় দাঁড়িয়ে প্রথমবার মাইক্রোফোন ধরেছিলাম পৃথিবীর জন্য ..
সত্যিই .. সময় কিভাবে চলে যায় !!



এগুলো যারা চেনেন তাদের শৈশব বড়ো মধুর ও আনন্দের ছিলো

Our Bengali Dataset

- The distribution of the processed Bengali dataset is shown below.
- This data is made keeping in mind the distribution of processed MVSA dataset.

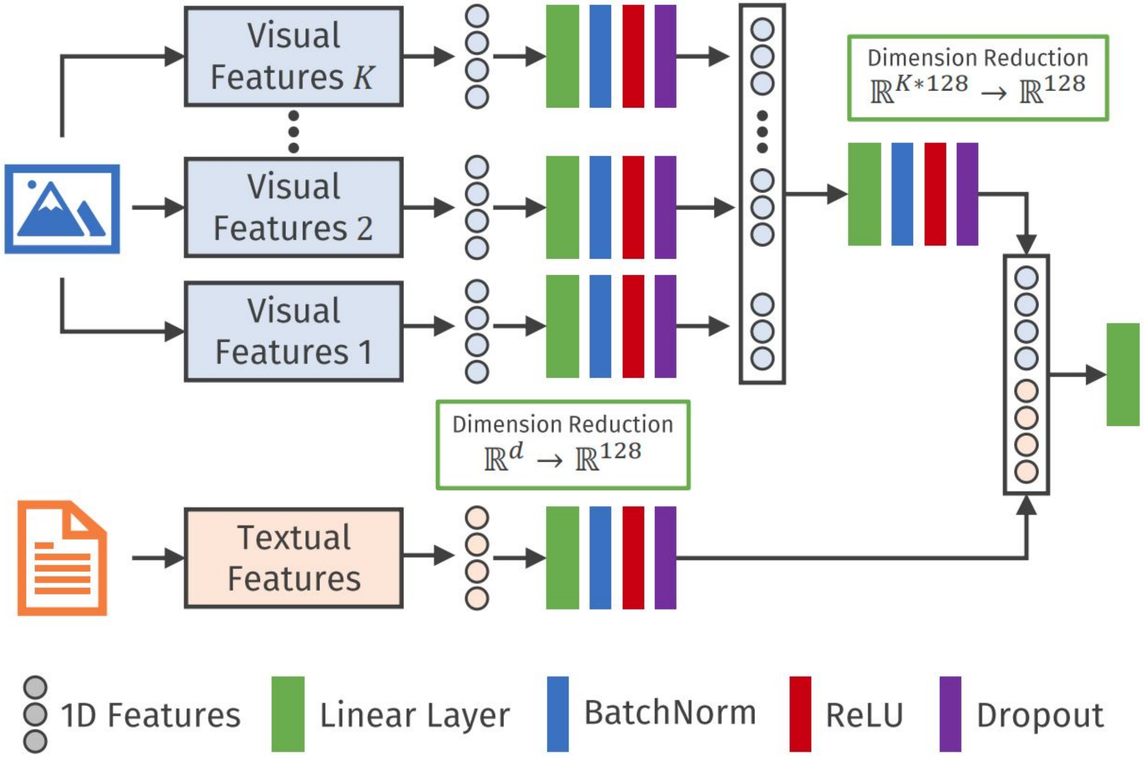


Models Used

- Model for Visual Feature Extraction:
  - Resnet101 trained on Imagenet dataset for object detection.
  - Resnet101 trained on Places365 dataset for scene detection.
  - Resnet50 trained on face images for facial expression detection.

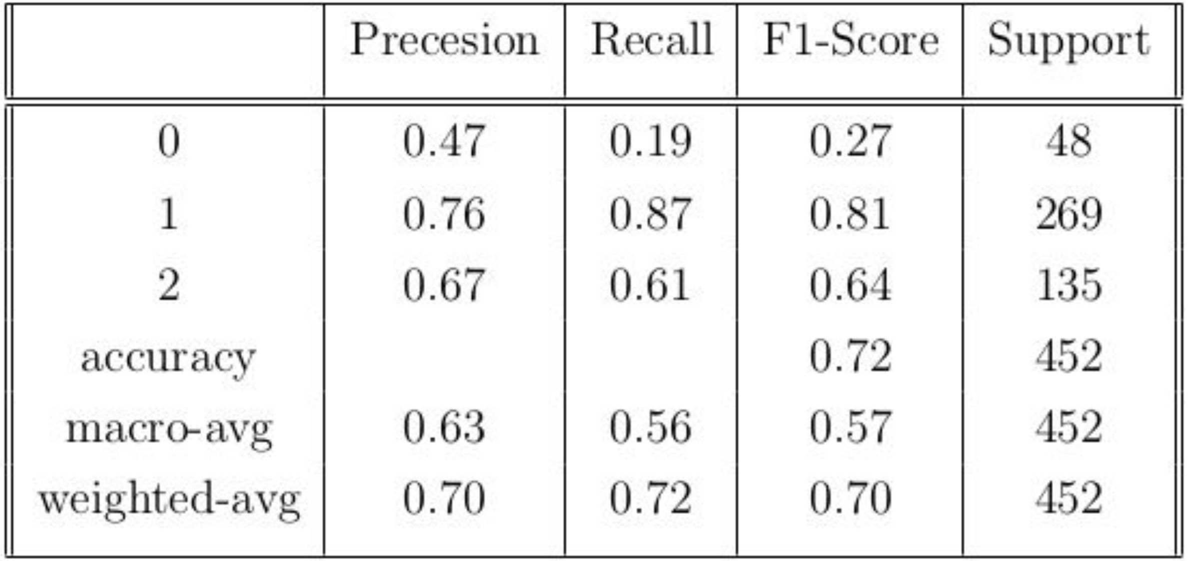
- Model for Textual Feature Extraction:
  - BERT-Cased model was used for extraction of textual features.
  - Or, Multilingual-BERT-Cased model was used when testing on Bengali dataset is done.

The Entire Model



Experimental Results

Validation Set Results for BERT-Cased Model

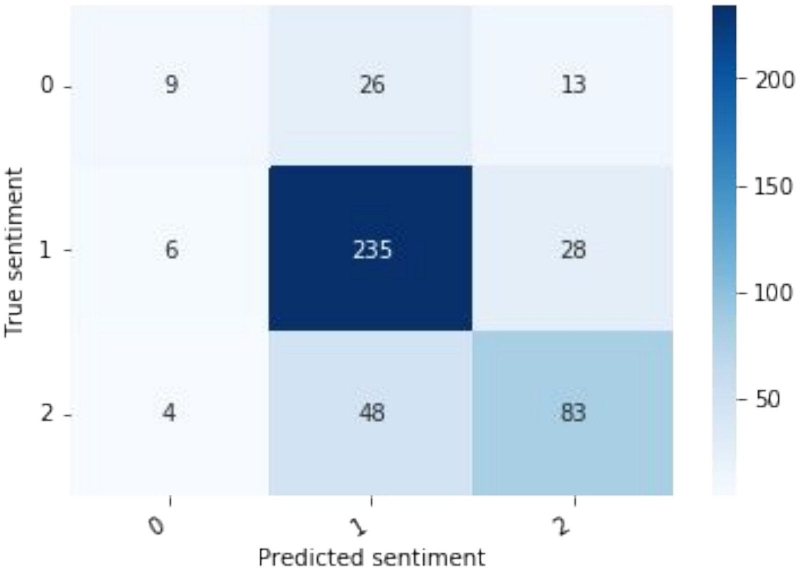


	Precesion	Recall	F1-Score	Support
0	0.47	0.19	0.27	48
1	0.76	0.87	0.81	269
2	0.67	0.61	0.64	135
accuracy			0.72	452
macro-avg	0.63	0.56	0.57	452
weighted-avg	0.70	0.72	0.70	452

Highest Accuracy = 72.34%

Experimental Results

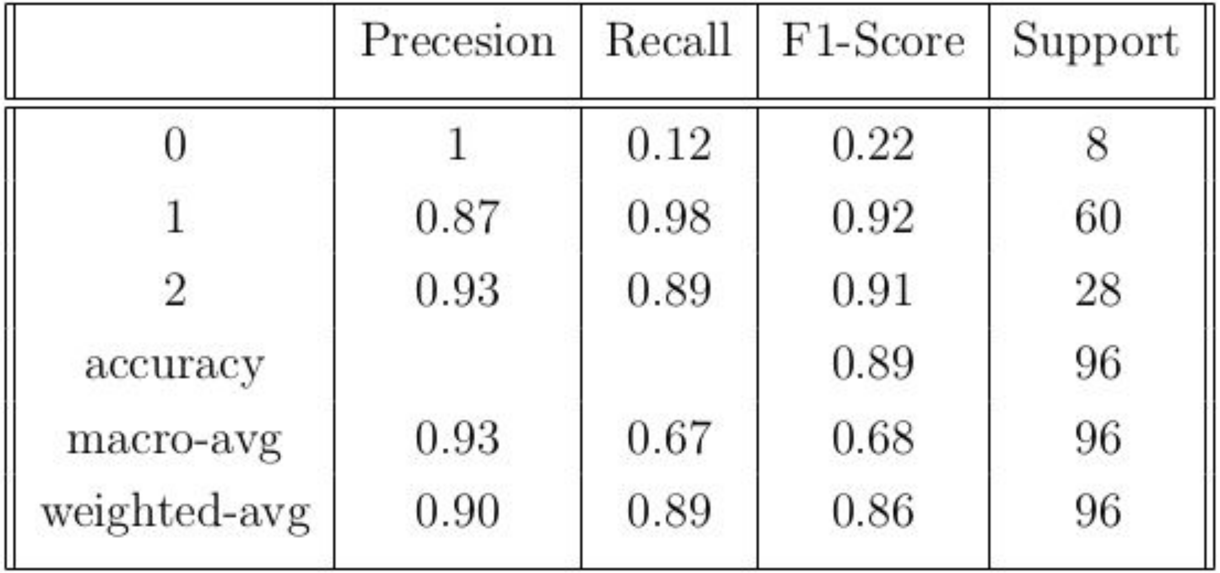
Confusion Matrix for **Validation** set using BERT-Cased Model



	0	1	2
0	9	26	13
1	6	235	28
2	4	48	83

Experimental Results

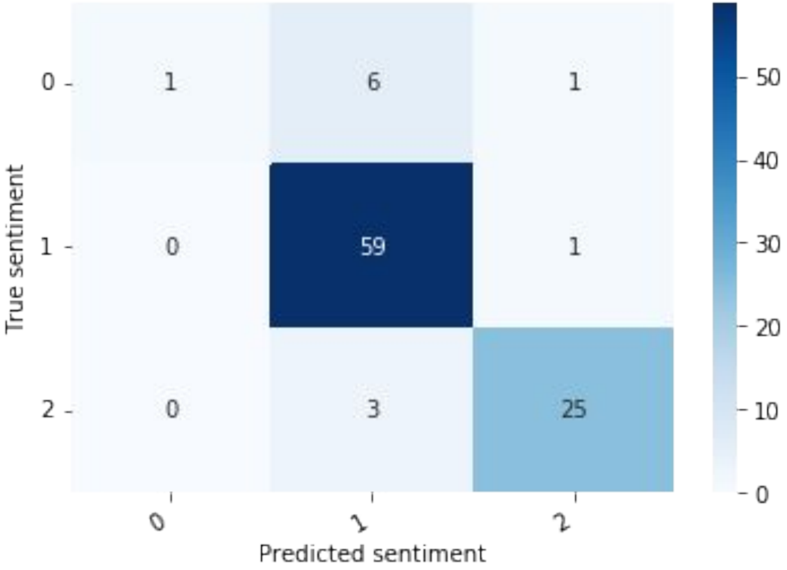
Metrics for **English** Test set using BERT-Cased Model



	Precesion	Recall	F1-Score	Support
0	1	0.12	0.22	8
1	0.87	0.98	0.92	60
2	0.93	0.89	0.91	28
accuracy			0.89	96
macro-avg	0.93	0.67	0.68	96
weighted-avg	0.90	0.89	0.86	96

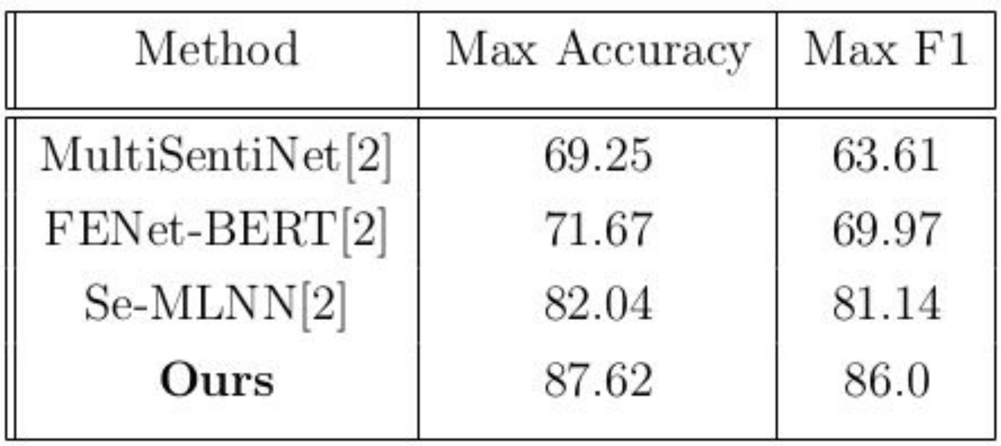
Experimental Results

Confusion Matrix for **English** Test set using BERT-Cased Model



Experimental Results

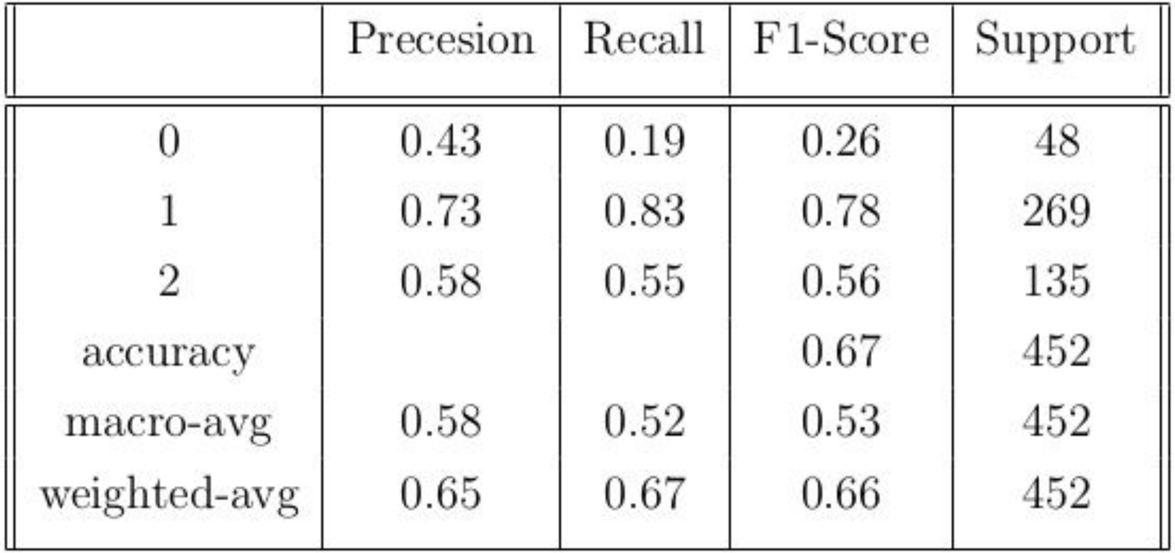
Comparison for **English** Test set using BERT-Cased Model with SOTA



Method	Max Accuracy	Max F1
MultiSentiNet[2]	69.25	63.61
FENet-BERT[2]	71.67	69.97
Se-MLNN[2]	82.04	81.14
Ours	87.62	86.0

Experimental Results

Validation Set Results for Multilingual-BERT-Cased Model

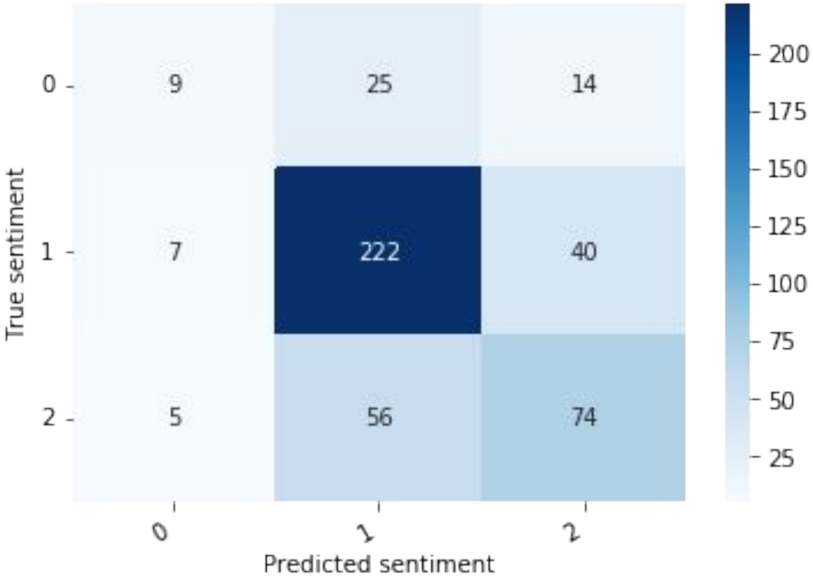


	Precesion	Recall	F1-Score	Support
0	0.43	0.19	0.26	48
1	0.73	0.83	0.78	269
2	0.58	0.55	0.56	135
accuracy			0.67	452
macro-avg	0.58	0.52	0.53	452
weighted-avg	0.65	0.67	0.66	452

Highest Accuracy = 67.47%

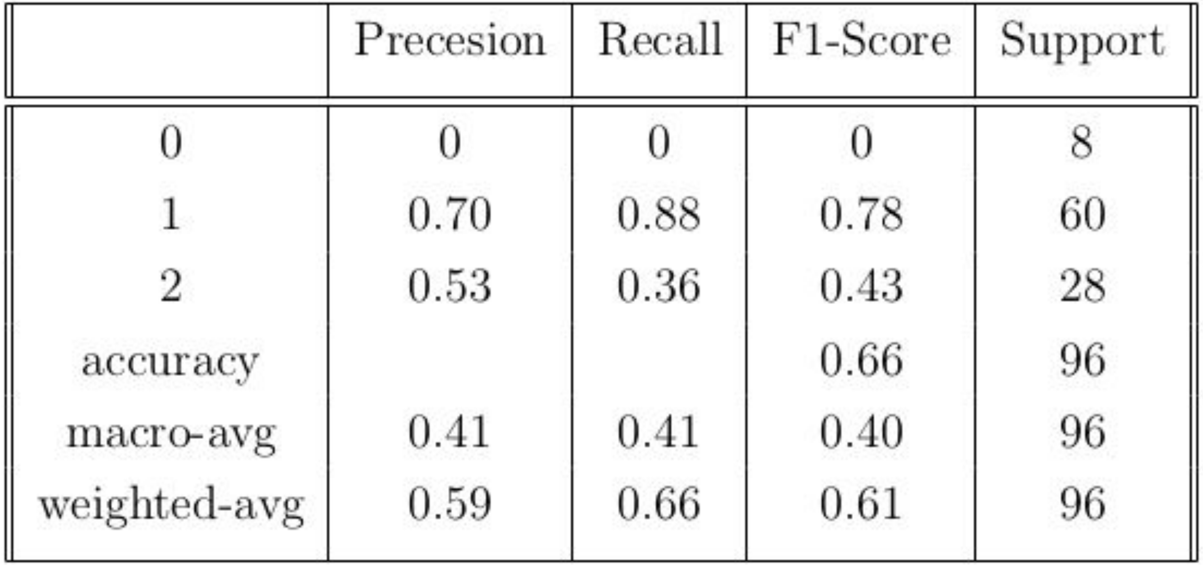
Experimental Results

Confusion Matrix for **Validation** set using Multilingual-BERT-Cased Model



Experimental Results

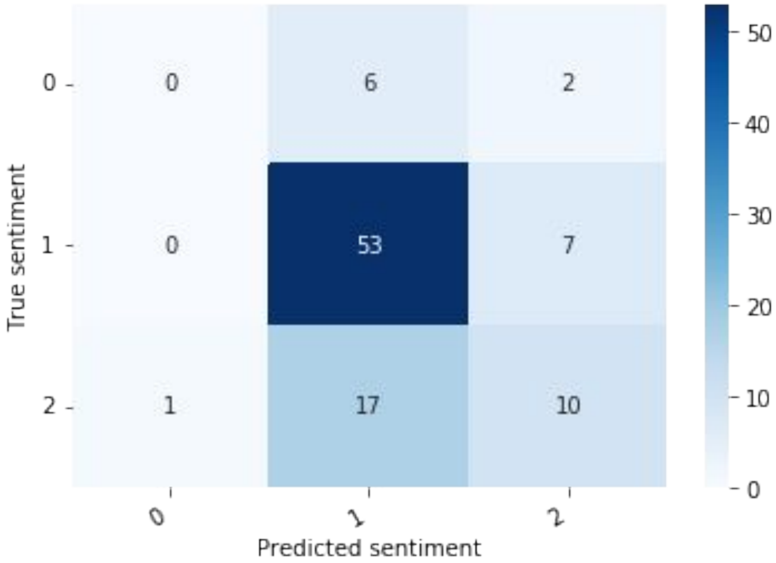
Metrics for **English** Test set using Multilingual-BERT-Cased Model



	Precesion	Recall	F1-Score	Support
0	0	0	0	8
1	0.70	0.88	0.78	60
2	0.53	0.36	0.43	28
accuracy			0.66	96
macro-avg	0.41	0.41	0.40	96
weighted-avg	0.59	0.66	0.61	96

Experimental Results

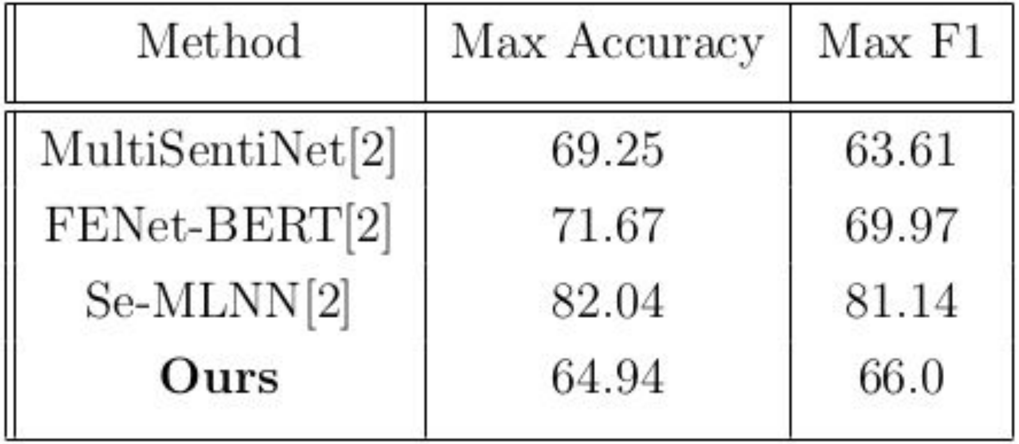
Confusion Matrix for **English** Test set using Multilingual-BERT-Cased Model



	0	1	2
0	0	6	2
1	0	53	7
2	1	17	10

Experimental Results

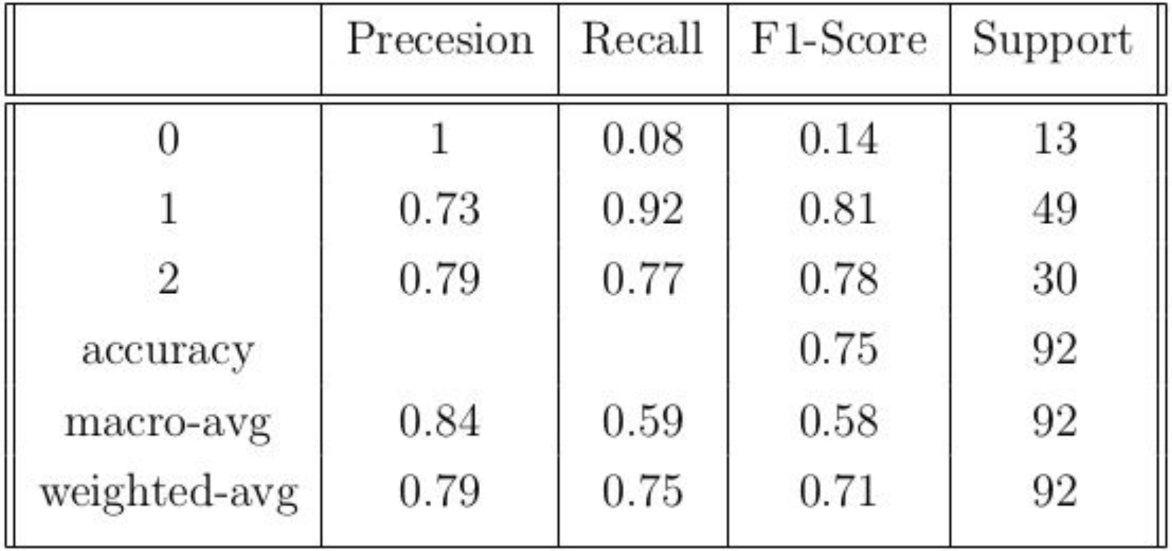
Comparison for **English** Test set using Multilingual-BERT-Cased Model with SOTA



Method	Max Accuracy	Max F1
MultiSentiNet[2]	69.25	63.61
FENet-BERT[2]	71.67	69.97
Se-MLNN[2]	82.04	81.14
Ours	64.94	66.0

Experimental Results

Metrics for **Bengali** Test set using Multilingual-BERT-Cased Model

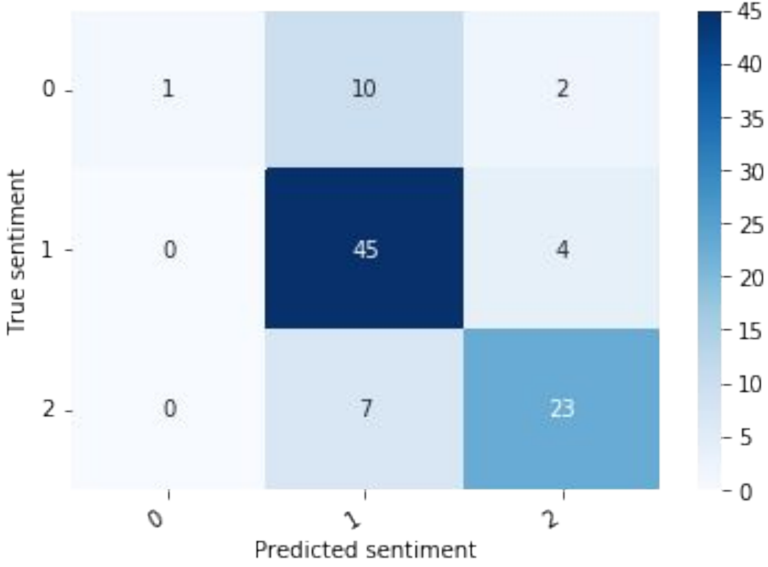


	Precesion	Recall	F1-Score	Support
0	1	0.08	0.14	13
1	0.73	0.92	0.81	49
2	0.79	0.77	0.78	30
accuracy			0.75	92
macro-avg	0.84	0.59	0.58	92
weighted-avg	0.79	0.75	0.71	92

Highest Accuracy = 71.13%

Experimental Results

Confusion Matrix for **Bengali** Test set using Multilingual-BERT-Cased Model



Summary

In summary, we have accomplished the followings:

- Experimentation with various image and encoder models like Resnet50, Resnet101, Bert-Cased, Multilingual-BERT-cased etc.
- Creation of an English language test dataset for evaluation
- Making our project truly multilingual by showing the capability of our model on Bengali test dataset.
- Achieved accuracy metrics close to the paper for the English test set using the BERT-cased model.

References

- **Dataset**: MULTIMEDIA COMMUNICATIONS RESEARCH LABORATORY. MVSA: Sentiment Analysis on Multi-view Social Data, In: https://mcrlab.net/research/mvsa-sentiment-analysis-on-multi-view-social-data/
- Paper Link: Gullal S. Cheema, Sherzod Hakimov, Eric Müller-Budack, and Ralph Ewerth, 2021, A Fair and Comprehensive Comparison of Multimodal Tweet Sentiment Analysis Methods. In Proceedings of the 2021 Workshop on Multi-Modal Pre-Training for Multimedia Understanding (MMPT '21). Association for Computing Machinery, New York, NY, USA, 37–45. https://doi.org/10.1145/3463945.3469058