> Link to the repository: REPO

Task 1

Scrape comments from the top 100 posts of the subreddit Sri Lanka: r/srilanka

Link to the Scrapping Code: /scrape.py

The scrapped Data is stored in this directory.

Scrapped Data: /comments

Task 2

Sentiment analysis from Hugging Face

> Models

- cardiffnlp/twitter-roberta-base-sentiment-latest
- finiteautomata/bertweet-base-sentiment-analysis
- Seethal/sentiment_analysis_generic_dataset

cardiffnlp : `21` Positive, `32` Negative, `48` Neutral %
finiteautomata : `6` Positive, `10` Negative, `84` Neutral %
Seethal : `4` Positive, `4` Negative, `92` Neutral %

> All three models show a heavy bias towards `neutral sentiments``. The "cardiffnlp" model has a more balanced distribution, while the "Seethal" and "finiteautomata" models is heavily skewed towards neutral predictions. The latter two models seem to have minimal positive and negative sentiment predictions.

`NOTE`: The output of the sentiment analysis is stored in the `out` folder. Later the combined output is stored in the `comments` directory. Please create `out` folder before running the code.

Link to Sentiment outputs: /comments

Link to Code: /sentiment_analysis

Task 3

Get the majority label by performing the majority vote

MAJORITY : `7` Positive, `10` Negative, `83` Neutral %

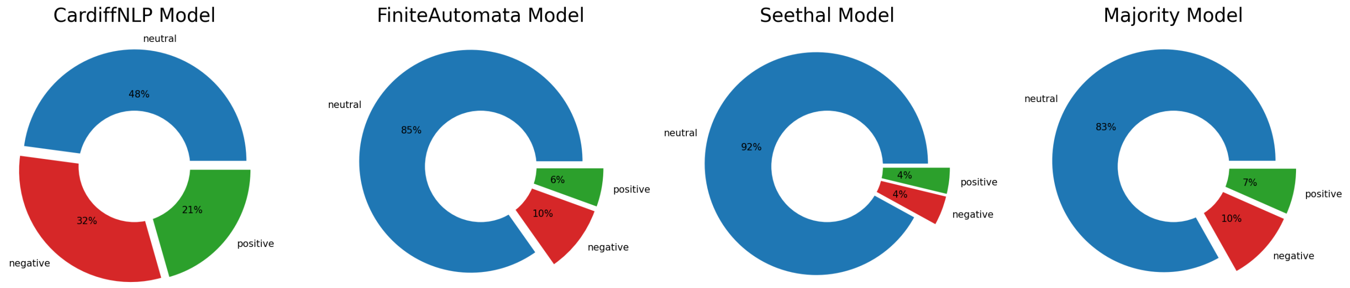

Fig: Sentiment Distribution of Majority Vote

Task 4

Do Exploratory Data Analysis (EDA)

Detailled results are stored in the EDA directory.

Link to COMMENT EDA: /EDA/EDA_Comments.ipynb

Link to POST EDA: /EDA/EDA_Posts.ipynb

COMMENTS:

1. Data Preprocessing for Exploratory Analysis

2. Comment Score Distribution and Comment Depth Distribution

3. Comment Depth vs Comment Score Relation

4. Distribution of Model Predictions for Sentiment Analysis

5. Distribution of Deleted Comments and Removed Accounts

5. Textual statistics

    - Number of characters.
    - Number of words.
    - Word length in each comment.

6. Stop Words v/s Non Stop Words

POST:

Titles of top 100 Reddit post from r/srilanka is taken.

1. Post Score Distribution:

2. Distribution of comments on Post

3. Post Type

4. Text statistics
    - Number of characters.

- Number of words.
- Word length in each comment.

5. Stop Words v/s Non Stop Words

Task 5

Randomly sample 100 sentences

Takes the combined dataset as the input. It contains all the comments from the 100 posts (5.7k) and samples 100 comments (34 positive, 33 negative, 33 neutral) and writes the output to a single file.

Link to the Sampling Code: **/sentiment_analysis/annotate_sampler.py**

Sampling is done as using following code snippet:

```
df_pos = df[df['MODEL_MAJORITY'] == 'positive'].sample(n=n_pos)
df_neg = df[df['MODEL_MAJORITY'] == 'negative'].sample(n=n_neg)
df_neu = df[df['MODEL_MAJORITY'] == 'neutral'].sample(n=n_neu)
```

Task 6

Human evaluation on these 100 sentences

Link to CSV annotation file: **/human_evaluation/human_eval.csv**

Task 7

The inter-annotator agreement using Krippendorff's alpha

Krippendorff's alpha (D) is a reliability metric used to assess agreement among multiple observers or methods in assigning distinctions to unstructured data, applicable across various fields. It measures data trustworthiness when multiple approaches generate data for analysis. [1]

Using the following formula;

$$\alpha = 1 - \frac{D_o}{D_e}$$

where D_o is the observed disagreement and D_e is the disagreement expected by chance.

Where D_e is calculated as follows;

$$D_e = \frac{1}{N(N-1)} \sum_{i=1}^{N} n_i \sum_{j=1}^{N} n_k \delta_{ij}^2$$

where delta_ij is the distance between the two annotations of the ith and jth items.

and D_o is calculated as follows;

$$D_o = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} O_{ij} \delta_{ij}^2$$

[1] Reference for the formula

Our results are as follows; Krippendorff's alpha: `0.3525931735261022`

The obtained Krippendorff's alpha value of `0.352` suggests that there exists some level of agreement among annotators, indicating a degree of shared understanding beyond random assignment of values. While not reaching ideal consensus (alpha=1), this value reflects a departure from the two extreme conditions—complete ambiguity and perfect agreement—highlighting the potential for improvement in achieving stronger inter-annotator consensus.

Link to the code: **/kripp.py**

Task 8

Perform a majority vote of the three annotators' labels

Link to the code: **/human_evaluation/model_annotate_mismatch.ipynb**

Link to the data: **/human_evaluation/human_eval.csv**

Task 9

Display 5 comments from the sample of 100 sentences where the majority label of models and majority label of human annotations is different,

COMMENT_ID	COMMENT_TEXT_CONTENT	ANNOTATORS_MAJORITY	MODEL_MAJORITY
j72bjki	I'd absolutely watch that!	positive	neutral

COMMENT_ID	COMMENT_TEXT_CONTENT	ANNOTATORS_MAJORITY	MODEL_MAJORITY
i4kvl63	Me, my husband and two boys are in SR right no...	neutral	positive
izsmqxw	Easiest geography map ever	positive	neutral
hajv6pv	Yeah those walksss! its a nice moment to keep ...	neutral	positive
gh9s8ju	Let's gooooo. I have no words thats really cool	neutral	positive

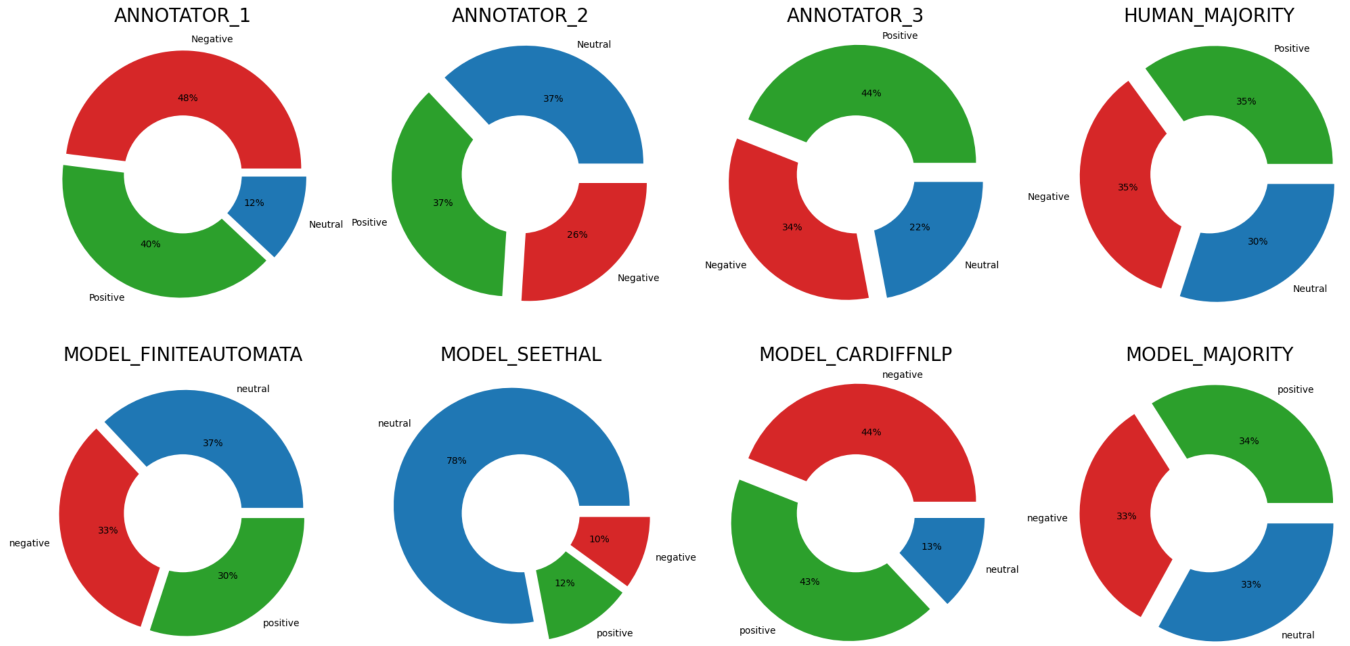

Fig: Comparision of annotators vs model prediction

*	POSITIVE	NEGATIVE	NEUTRAL
ANNO_1 (22210019)	40	48	12
ANNO_2 (22210034)	37	26	37
ANNO_3 (22210036)	44	34	22
CARDIFFNLP	43	44	13
FINITEAUTOMATA	30	33	37
SEETHAL	12	10	78
ANNO_MAJORITY	35	35	30
MODEL_MAJORITY	34	33	33

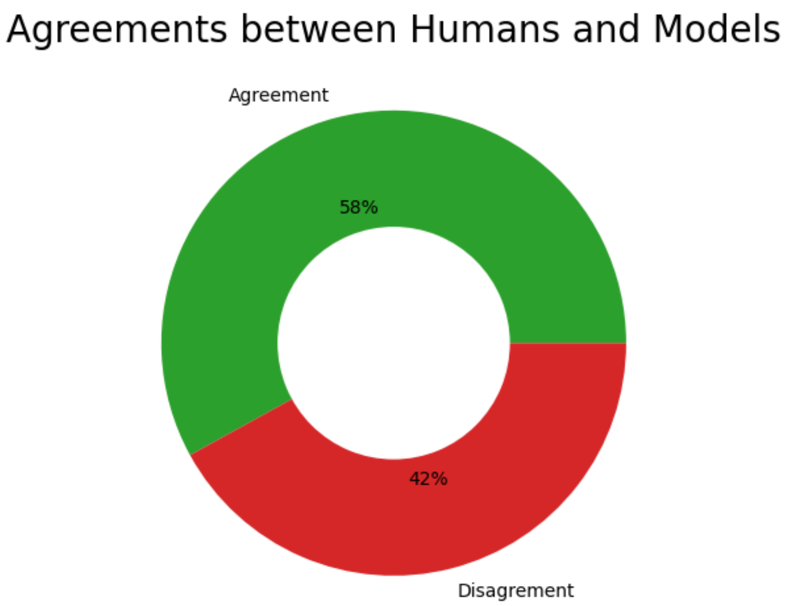

Fig: Agreement of annotators vs model prediction

Link to the code: **/human_evaluation/model_annotate_mismatch.ipynb**

Result 1

EDA and Human evaluation with respect to sentiment analysis.

> All three models show a heavy bias towards `neutral sentiments`. The "cardiffnlp" model has a more balanced distribution, while the "Seethal" and "finiteautomata" models is heavily skewed towards neutral predictions. The latter two models seem to have minimal positive and negative sentiment predictions.

> Models

- cardiffnlp/twitter-roberta-base-sentiment-latest
- finiteautomata/bertweet-base-sentiment-analysis
- Seethal/sentiment_analysis_generic_dataset

> Sentiment Distribution of each model

cardiffnlp : `21` Positive, `32` Negative, `48` Neutral %

> *This model seems to produce a relatively balanced distribution of sentiments, with a higher percentage of neutral predictions.*

finiteautomata : `6` Positive, `10` Negative, `84` Neutral %

> *This model appears to predict a majority of neutral sentiments, with a small portion of positive and negative sentiments*

Seethal : `4` Positive, `4` Negative, `92` Neutral %

> *This model predominantly predicts neutral sentiments, with very low percentages of positive and negative sentiments*

MODEL MAJORITY : `7` Positive, `10` Negative, `83` Neutral %

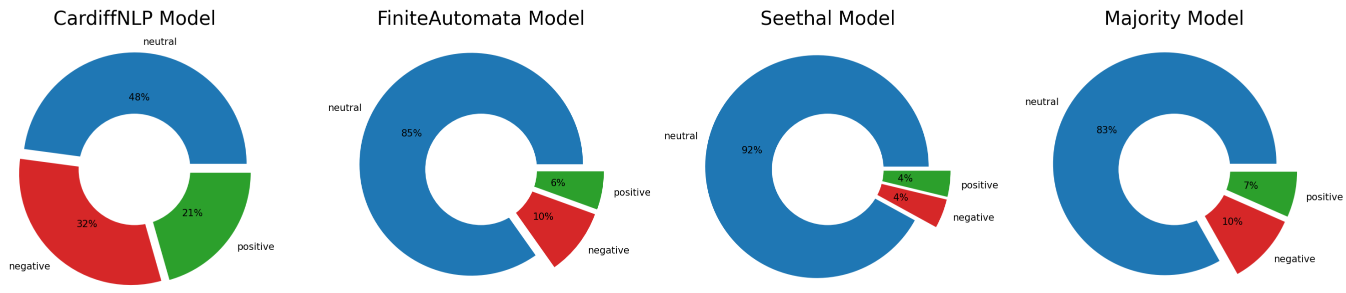

Fig: Sentiment Distribution of MODEL's predictions

Sentiment Distribution of each Annotator

Annotator 1 : `40` Positive, `48` Negative, `12` Neutral %

> Annotator 1 provided a sentiment distribution with 40% positive, 48% negative, and 12% neutral sentiments.

Annotator 2 : `37` Positive, `26` Negative, `37` Neutral %

> Annotator 2's sentiment distribution consisted of 37% positive, 26% negative, and 37% neutral sentiments.

Annotator 3 : `44` Positive, `34` Negative, `22` Neutral %

> Annotator 3's sentiment distribution indicated 44% positive, 34% negative, and 22% neutral sentiments.

Annotator MAJORITY : `35` Positive, `35` Negative, `30` Neutral %

> The majority sentiment distribution from annotators revealed 35% positive, 35% negative, and 30% neutral sentiments.

Human annotators display varying degrees of sentiment distribution, with Annotator 1 showing a notable negative bias, Annotator 2 having a more balanced approach, and Annotator 3 slightly favoring positive sentiments. The majority sentiment distribution among the annotators reveals a more even distribution of sentiments, indicating a collective understanding of the sentiment spectrum.

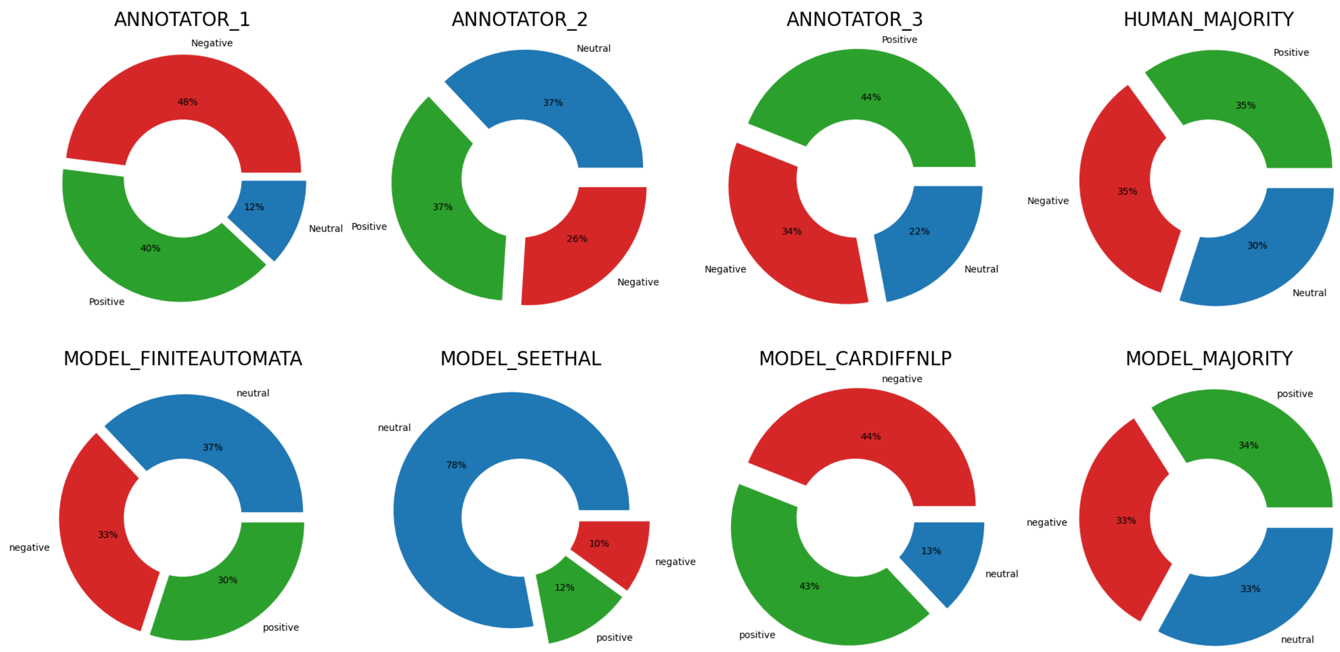

Fig: Sentiment Distribution of Annotator's predictions

> The models' strong bias towards neutral predictions suggests a need for further refinement to accurately capture positive and negative sentiments present in the data. The variation in human annotators' results indicates the subjectivity of sentiment analysis and highlights the importance of considering diverse perspectives in evaluating sentiment.

The `cardiffnlp` model's sentiment distribution shows a closer match to the majority annotators' sentiment distribution, particularly in the balance between positive, negative, and neutral sentiments.

Model v/s Annotator Agreement

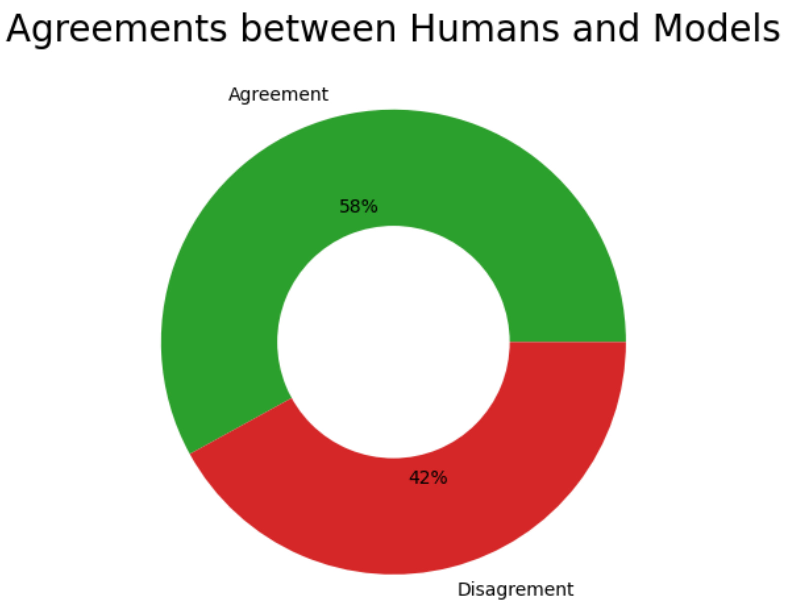

Fig: Agreement of annotators vs model prediction

Matched Labels : `58`

Unmatched Labels : `42`

Result 2

Link to the code: [**/word_cloud/word_cloud.py**](/word_cloud/word_cloud.py)

.

.

Task Contribution

Sl. No	Name	Roll No.	Tasks	Contribution %
1	Ritesh Patidar	22210034	6, 8, 9	11
2	Siddhesh Dosi	22210045	2, 3, 5, Result-1	11
3	Anupam Sharma	22210006	1, Result-1	11
4	Kowsik Nandagopan D	22250016	2, 3, 5	12
5	Ankit Yadav	22270001	4, Result-1	11
6	Sai Krishna Avula	22210036	6, 8, 9	11
7	Hitesh lodwal	22210019	6, 8, 9	11
8	Madhuri Awachar	22210023	7, Result-2	11
9	Ayush Shrivastava	22210010	4, 7	11