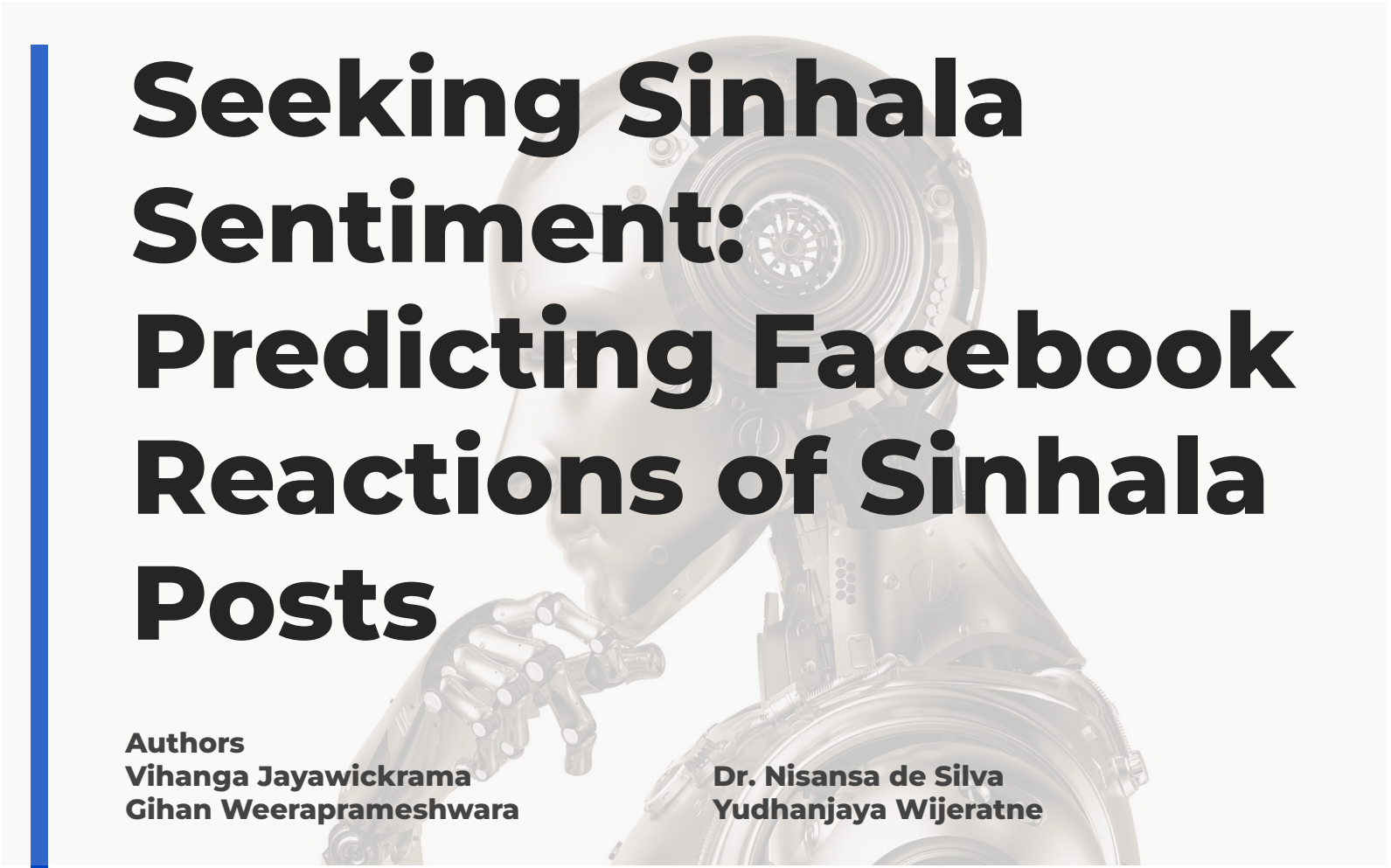
# Seeking Sinhala Sentiment: Predicting Facebook Reactions of Sinhala Posts



**Authors**
**Vihanga Jayawickrama**
**Gihan Weeraprameshwara**

**Dr. Nisansa de Silva**
**Yudhanjaya Wijeratne**

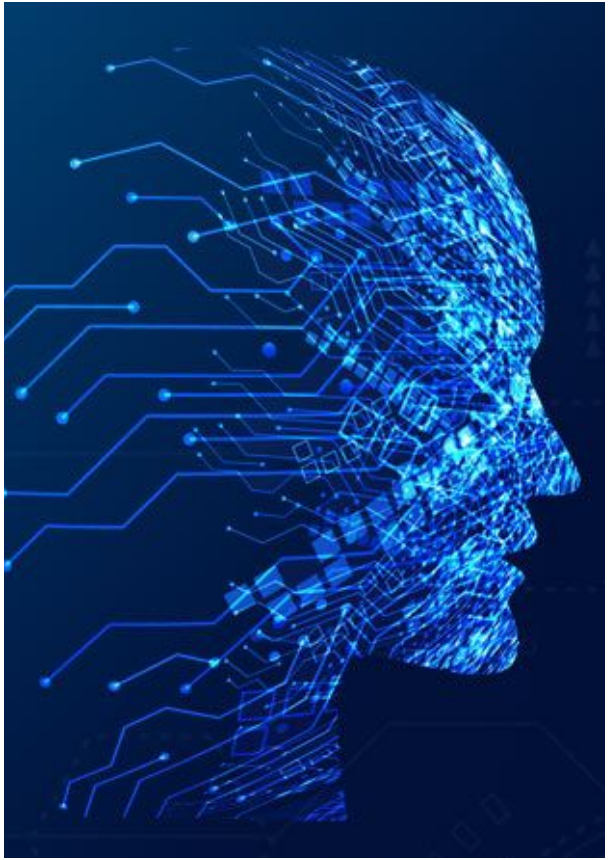
Looking at things from other people's point of view is practically the secret of success."

—Paul Graham

# Outline

1. Introduction
2. Related Work
3. Dataset
4. Methodology
5. Outcomes

# 1. Introduction



# Sentiment Analysis Models for Sinhala Colloquial Text

### Reader's Perspective Instead of the Writer's Perspective

Product reviews in E-commerce sites, movie review analysis, political analysis use the writer's perspective

### Predicts the Facebook Reactions, not the Real Sentiment

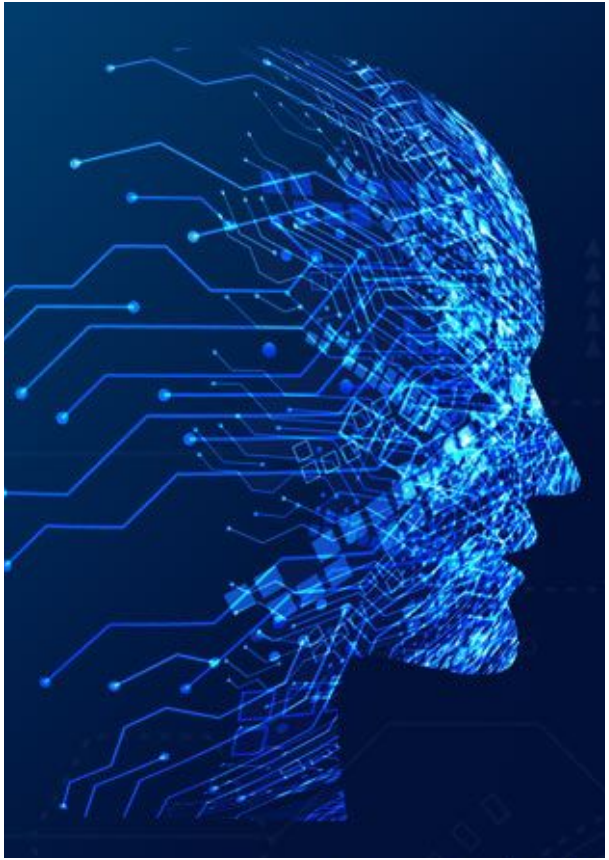
Different people associate different sentiments with the same Facebook reaction

# Steps

- **Data pre processing**

- **Creating three baseline models**
  - All Reaction Set Model - All reactions
  - Core Reaction Set Model - Excluded Like and Thankful
  - Star Rating Model - Love, Wow as positive;  Sad, Angry as negative
- **Testing the dataset**

# 2. Motivation



- **Tests the usability of Facebook data as a convenient means for sentiment analysis**
- **The first attempt to create a sentiment analysis model through the reader’s perspective for Sinhala**
- **Enables Facebook users to estimate the reception of their posts prior to posting**
- **Contributes towards eradicating the research poverty of the Sinhala language**
- **Provides the means and methods to continue research in NLP on Sinhala language**

# Potential Use Cases

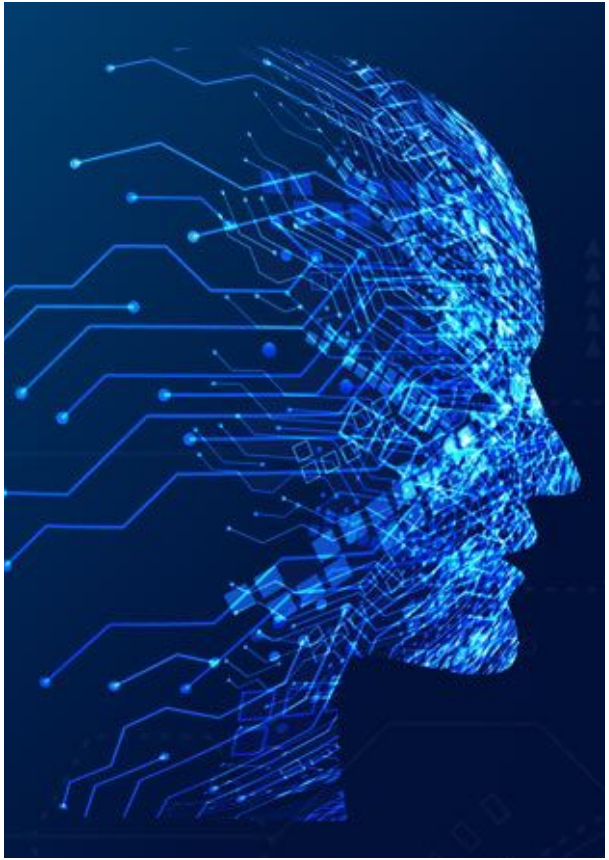
## Brand Management

Provides the ability to understand how the audience would perceive a Facebook post before posting it

## Support Researches Based on Social Networks

The outcomes can be used to understand how the Sri Lankan crowd respond to the current social media content.
This research can build a cornerstone for more advanced tools for Sinhala language.

# 2. Related Work



# Related Work

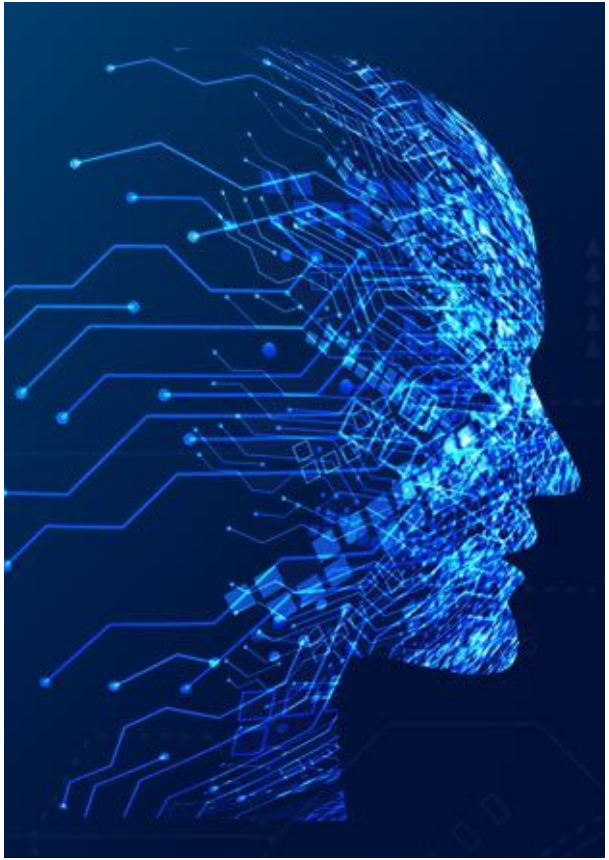
- **Sentiment analysis**
  - Negativity of parliament speeches by Rudkowsky et al. (2018)
  - Identify real-world events using Population Sentiment Orientation of social media data by L Che
  - A survey of sentiment analysis in social media by L. Yue et al. (2019)

- **Facebook data related sentiment analysis research**
  - Understanding facebook reactions to scholarly articles by Freeman et al. (2019)
  - Work of Measuring the diversity of facebook reactions to research by Freeman et al. (2020)
  - Facebook sentiment: Reactions and emojis by Y.Tian et al.(2017)

# Sinhala NLP research

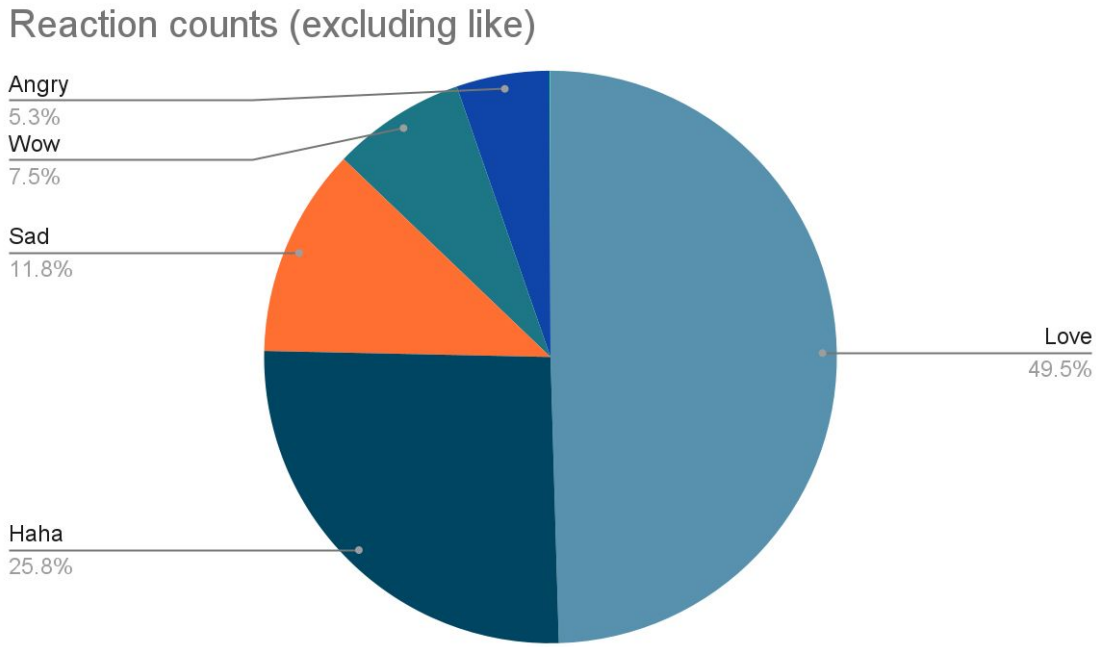
- **Survey on publicly available sinhala natural language processing tools and research by N. de Silva (2019)**
- **Sentiment analysis for sinhala language using deep learning techniques by Senevirathne et al. (2020)**
- **Sinhala language corpora and stopwords from a decade of sri lankan facebook by Wijeratne and N. de Silva (2020)**

# 3. Dataset



# Dataset Used

- Developed by Mr. Yudhanjaya Wijeratne and Dr. Nisansa de Silva.
- Contains 1.8 million Facebook posts spanning over a decade from different sources.

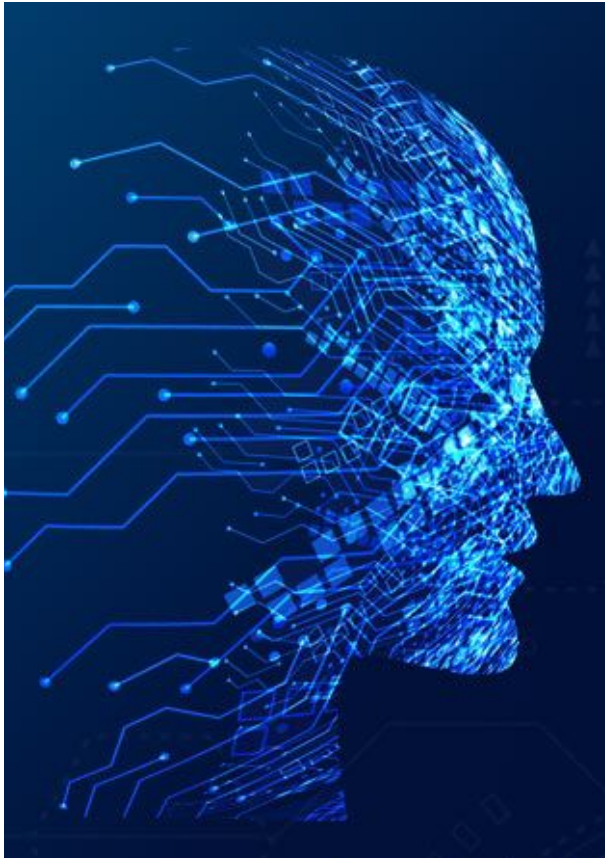


- Over 540 million user reactions
- 526,732 data rows after preprocessing steps

| FIELD | DATA TYPE | DESCRIPTION |
|---|---|---|
| Index | Integer | Index of the entry |
| Count of each reaction* | Integer | The count of each reaction given to each post |
| Message | String | Textual content of the Facebook post |

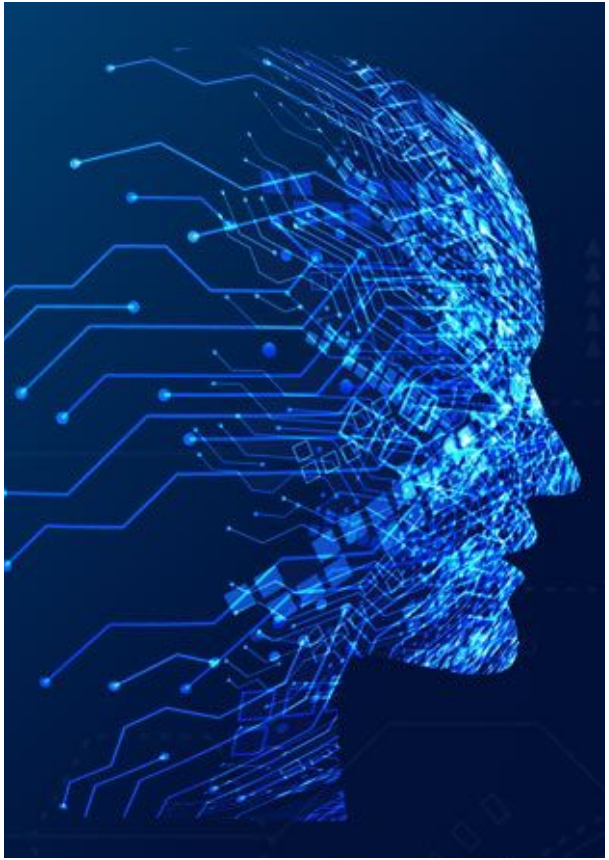
*Like, love, wow, haha, sad, angry, thankful

# 4. Objectives

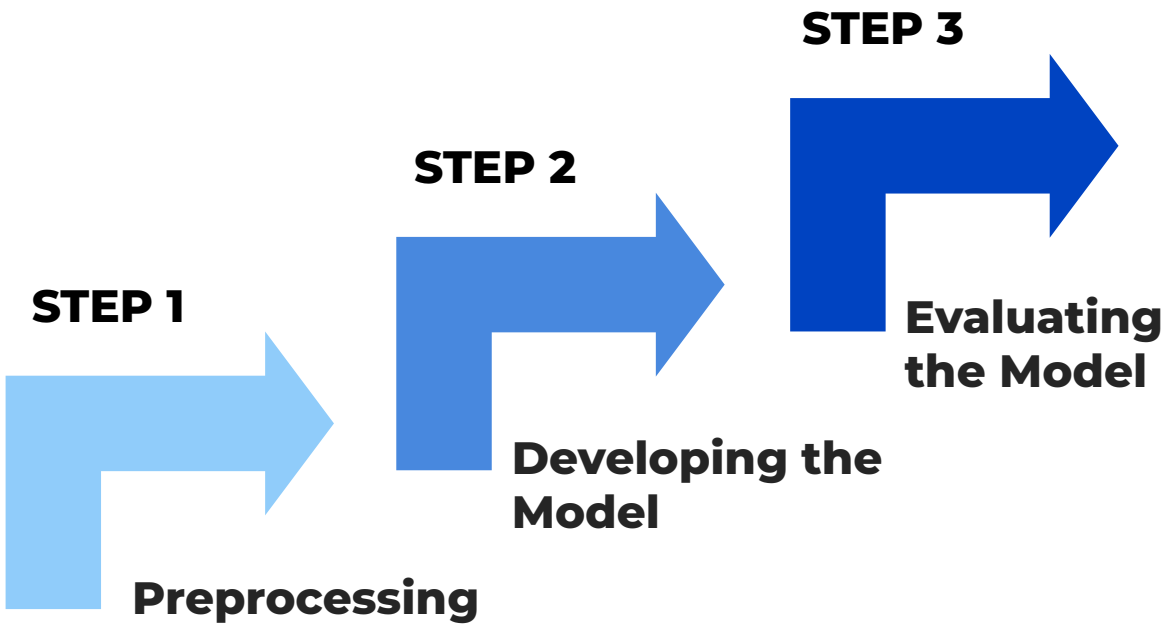


- **Develop a tool to detect sentiments in colloquial Sinhala text**
- **Develop a tool to predict the Facebook user reactions to Sinhala text.**
- **Compare with current tools to determine the best option.**
- **Test the effectiveness of Facebook data for Sinhala language based sentiment analysis.**
- **Introduce a methodology to develop accurate Sinhala NLP tools.**
- **Provide the means and methods to continue research in NLP on Sinhala language.**

# 4. Methodology



# Walkthrough



**1.**
**Text Cleaning and Normalizing Reactions**

- Removing text in other languages, numbers, and other text that contains no sentimental value
- Scaling the reaction counts of each row so that their sum is 1
- Removes the bias towards posts with higher reaction counts

**2.**
**Tokenizing Words**

- Dividing each message into word tokens and removing stopwords
- Stopwords do not contain a significant sentimental value

**3.**

**Vectorizing Sentiments**

- Assigning the normalised reaction vector of a post to each of its word tokens
- For words included in multiple posts, the mean value of vectors assigned to them is taken

**4.**

**Training and Testing**

- Selecting the correct models and techniques
- Hyper parameter tuning
- Training  and testing

# Core Reaction Set Model

- **A model to predict the proportions of Love, Wow, Haha, Sad, and Angry reactions**

- **Excludes Like and Thankful**

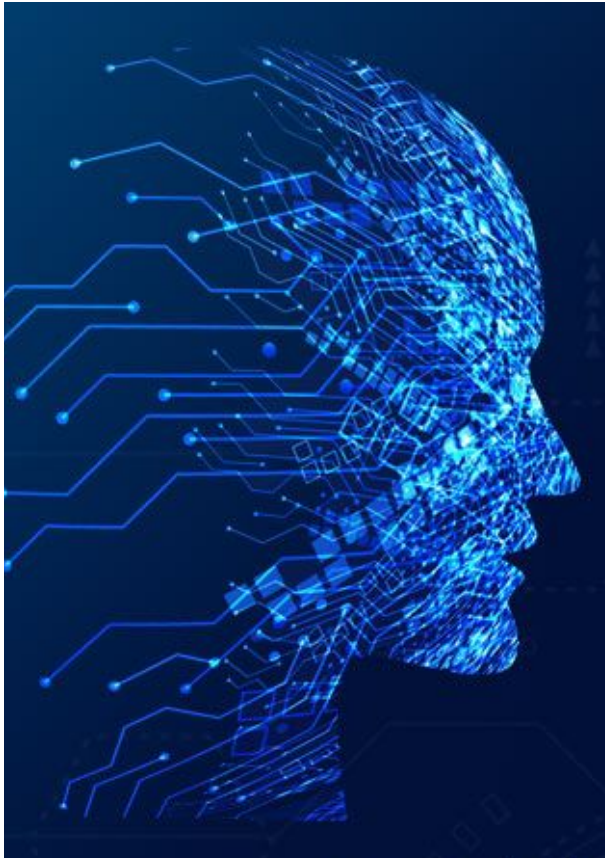
# All Reaction Set Model

- **Considers Facebook reactions Like, Love, Wow, Haha, Sad, Angry, and Thankful**
- **High bias towards Like and struggles to identify other reactions**

# Star Rating Model

- **Uses the star rating schema to quantify sentiment**
- **Love and Wow is used as positive sentiment; Angry and Sad is used as negative sentiment; Haha is neglected.**

# 5. Outcomes



# Results: Core Reactions vs. All Reactions

| Train(%) | Reaction | Maximum F1 Score Achieved | |
|---|---|---|---|
| | | Core Reaction | All Reaction |
| 95% | Like | - | 0.9626 |
| | Love | 0.5164 | 0.1769 |
| | Wow | 0.2218 | 0.0818 |
| | Haha | 0.3060 | 0.1068 |
| | Sad | 0.1613 | 0.0638 |
| | Angry | 0.1318 | 0.0495 |
| | Thankful | - | 0.0000 |

# Results: Star Rating

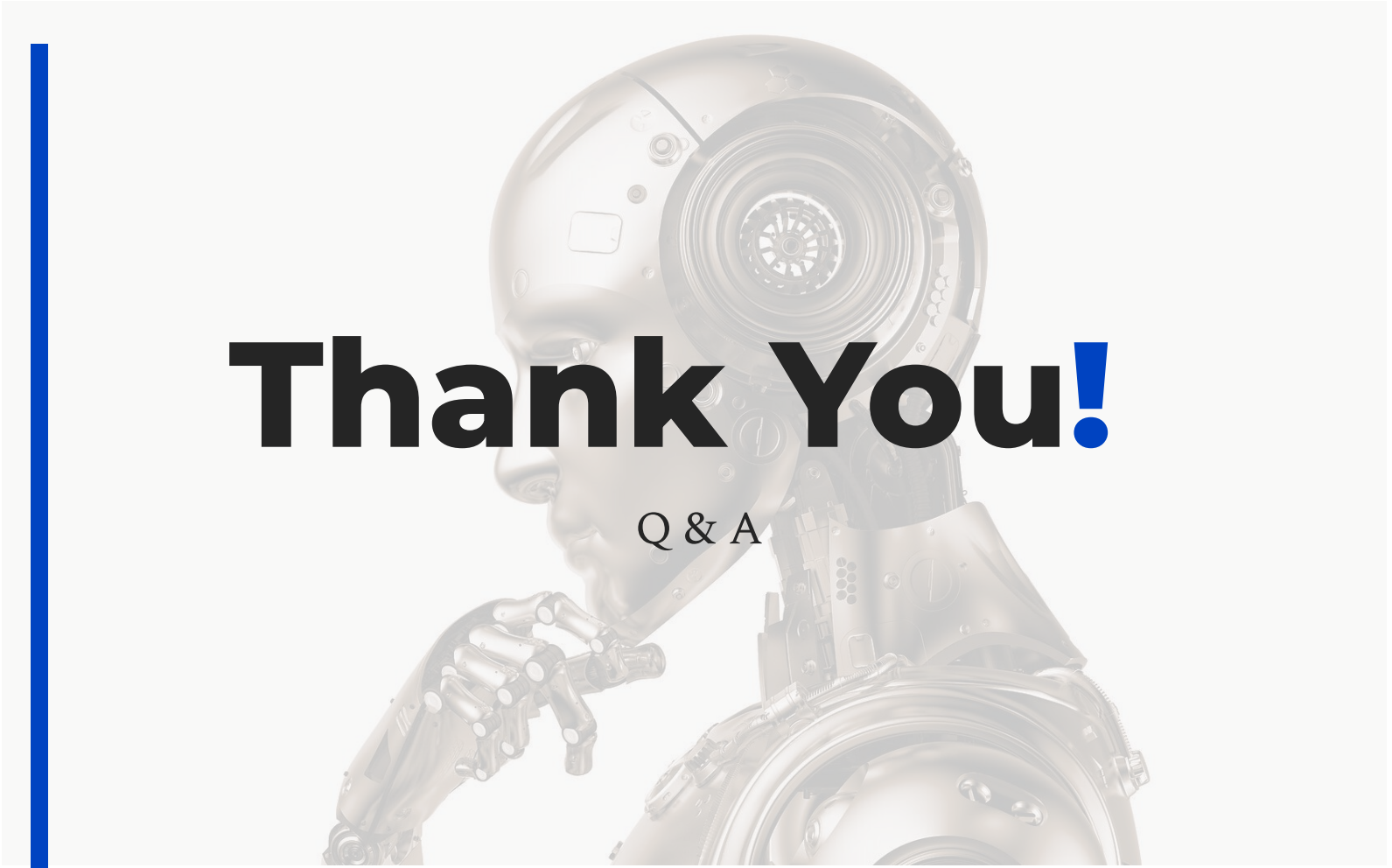
| Train (%) | Category | Maximum F1 score Achieved |
|---|---|---|
| 95% | Positive | 0.7068 |
| | Negative | 0.4207 |
| | Star Rating | 0.2921 |

- **Inclusion of Like reaction hinders the ability of the model to accurately predict other reactions.**
- **Aggregation of reactions into a single Star Rating value has caused a decrease in F1 Score.**
- **Performance of models is largely unaffected by the train-test division.**

# References


- E. Rudkowsky, M. Haselmayer, M. Wastian, M. Jenny,ˇS. Emrich, and M. Sedlmair, “More than bags of words: Sentiment analysis with word embeddings,” Communication Methods and Measures, vol. 12, no.2-3, pp. 140–157, 2018.
- L. Che, “Sentiment-based spatial-temporal event detection in social media data.”
- L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, “A survey of sentiment analysis in social media,”Knowledge and Information Systems, vol. 60, no. 2, pp. 617–663, 2019.
- C. Freeman, M. K. Roy, M. Fattoruso, and H. Alhoori,“Shared feelings: Understanding facebook reactions to scholarly articles,” in 2019 ACM/IEEE JCDL.IEEE, 2019, pp.301–304.
- C. Freeman, H. Alhoori, and M. Shahzad, “Measuring the diversity of facebook reactions to research,” Proceedings of the ACM on Human-Computer Interaction,vol. 4, no. GROUP, pp. 1–17, 2020
- Y. Tian, T. Galery, G. Dulcinati, E. Molimpakis, and C. Sun, “Facebook sentiment: Reactions and emojis,”in Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media. Valencia, Spain: Association for Computational Linguistics, Apr. 2017, pp. 11–16. [Online]. Available:https://www.aclweb.org/anthology/W17-1102

- C. Pool and M. Nissim, “Distant supervision for emotion detection using facebook reactions,”arXiv preprint arXiv:1611.02988, 2016
- L. Graziani, S. Melacci, and M. Gori, “Jointly learning to detect emotions and predict facebook reactions,”in International Conference on Artificial Neural Networks. Springer, 2019, pp. 185–197.
- F. Krebs, B. Lubascher, T. Moers, P. Schaap, and G. Spanakis, “Social emotion mining techniques for facebook posts reaction prediction,”arXiv preprint arXiv:1712.03249, 2017
- N. de Silva, “Survey on publicly available sinhala natural language processing tools and research,”arXiv preprint arXiv:1906.02358, 2019.
- N. Medagoda, S. Shanmuganathan, and J. Whalley,“Sentiment lexicon construction using sentiwordnet3.0,” in 2015 11th International Conference on Natural Computation (ICNC). IEEE, 2015, pp. 802–807.
- L. Senevirathne, P. Demotte, B. Karunanayake, U. Munasinghe, and S. Ranathunga, “Sentiment analysis for sinhala language using deep learning techniques,” 2020.
- Y. Wijeratne and N. de Silva, “Sinhala language corpora and stopwords from a decade of sri lankan facebook,”arXiv preprint arXiv:2007.07884, 2020.



# Thank You!

Q & A