**A Framework for Social Media Opinion Mining for Low Resource Marathi Text**

**Dataset Pre-processed and ready for implementation:**

Data before:

* Contained text in English, mentions, hashtags, emojis and links
* We removed all the mentions, emojis and links.
* The English hashtags were totally removed while the hashtags in marathi were kept.
* Marathi words between the text that are hashtag can have an influence and be useful in the tweet.

Text

Description automatically generated

After:

* Only marathi text left, for the implementation

Timeline

Description automatically generated with medium confidence

* Started understanding the XLM-R architecture.
* Will use transformer XLM-R with our processed tweets dataset for fine-tuning the language model.
* Will train and compare results with models like mBERT and BiLSTM for accuracy as previous papers showed our proposed model performing better than the others for multilingual tasks.
* Start with the framework