

# Overview of the track on Sentiment Analysis for Dravidian Languages in Code-Mixed Text

Bharathi Raja Chakravarthi<sup>a</sup>, Ruba Priyadharshini<sup>b</sup>, Vigneshwaran Muralidaran<sup>c</sup>,  
Shardul Suryawanshi<sup>a</sup>, Navya Jose<sup>d</sup>, Elizabeth Sherly<sup>d</sup> and John P. McCrae<sup>a</sup>

<sup>a</sup>Insight SFI Research Centre for Data Analytics, Data Science Institute, National University of Ireland Galway.

<sup>b</sup>ULTRA Arts and Science College, Madurai, Tamil Nadu, India

<sup>c</sup>School of Computer Science and Informatics, Cardiff University, United Kingdom

<sup>d</sup>Indian Institute of Information Technology and Management-Kerala, India

## Abstract

Sentiment analysis of Dravidian languages has received attention in recent years. However, most social media text is code-mixed, and there is no research available on the sentiment analysis of code-mixed Dravidian languages. The Dravidian-CodeMix-FIRE 2020 <https://dravidian-codemix.github.io/2020/>, a track on Sentiment Analysis for Dravidian Languages in Code-Mixed Text, focused on creating a platform for researchers to come together and investigate the problem. Two language tracks, Tamil and Malayalam, were created as a part of Dravidian-CodeMix-FIRE 2020. The goal of this shared task was to identify the sentiment of a given code-mixed comment (from YouTube) into five classes - *positive*, *negative*, *neutral*, *mixed-feeling* and *comment not in the intended language*. The performance of the systems (developed by participants) has been evaluated in terms of weighted-F1 score.

## Keywords

sentiment analysis, Dravidian languages, Tamil, Malayalam, code-mixing, text classification, deep learning

## 1. Introduction

Sentiment analysis is the task of identifying subjective opinions or responses about a given topic. Sentiment analysis on social media reveals to you how individuals feel about your brand on the web. As opposed to a basic check of notices or remarks, supposition examination thinks about feelings and assessments. It includes gathering and breaking down data in the posts individuals share about your brand via online media. It has been an active area of research in the past two decades in both academia and industry. There is an increasing demand for sentiment analysis on social media texts which are largely code-mixed. Code-mixing is a prevalent phenomenon in a multilingual community where the words, morphemes and phrases from two or more languages are mixed in speech or writing [1]. A few researchers utilize the expressions "code-mixing" and "code-switching" reciprocally, particularly in investigations of linguistic structure, morphology, and other proper parts of language. Code-mixed texts are often written in non-native scripts particularly on social media [2]. Hence, systems trained on monolingual data fail on code-mixed data due to the complexity introduced by code-switching at different linguistic levels in the text [3]. This shared task presents a new gold standard corpus

---

*FIRE 2020: Forum for Information Retrieval Evaluation, December 16-20, 2020, Hyderabad, India*

EMAIL: bharathi.raja@insight-centre.org (B.R. Chakravarthi); rubapriyadharshini.a@gmail.com (R. Priyadharshini);

vigneshwar.18@gmail.com (V. Muralidaran); shardul.suryawanshi@insight-centre.org (S. Suryawanshi);

navya.mi3@iiitmk.ac.in (N. Jose); sherly@iiitmk.ac.in (E. Sherly); john.mccrae@insight-centre.org (J.P. McCrae)

ORCID: 0000-0002-4575-7934 (B.R. Chakravarthi)



© 2020 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

for sentiment analysis of code-mixed text in Dravidian languages (Tamil-English and Malayalam-English).

Tamil is one of the Dravidian languages spoken by Tamil people in India, Sri Lanka and by the Tamil diaspora around the world, with official recognition in India, Sri Lanka and Singapore. Malayalam is another Dravidian language spoken in the southern region of India with official recognition in the Indian state of Kerala and the Union Territories of Lakshadweep and Puducherry [4, 5]. There are nearly 75 million Tamil speakers<sup>1</sup> and 45 million Malayalam speakers<sup>2</sup> in India and other countries. Tamil and Malayalam are highly agglutinative languages [6, 7].

Tamil script evolved from the Tamil script<sup>3</sup>, Vatteluttu alphabet, and Chola-Pallava script. The modern Tamil script descended from the Chola-Pallava script. It has 12 vowels, 18 consonants, and 1 āydam (voiceless velar fricative). Minority languages such as Saurashtra, Badaga, Irula, and Paniya are also written in the Tamil script [8, 9]. The Malayalam script is the Vatteluttu alphabet extended with symbols from the Grantha script. Both Tamil and Malayalam scripts are alpha-syllabic, belonging to a family of the abugida writing systems that are partially alphabetic and partially syllable-based [10, 11, 12]. However, social media users often adopt Roman script for typing as it is easy to input. Hence, the majority of the data available in social media for these under-resourced languages are code-mixed.

The goal of this task is to identify the sentiment polarity of the code-mixed dataset of comments/posts in Dravidian Languages (Malayalam-English and Tamil-English) collected from social media. The comment/post may contain more than one sentence but the average sentence length of the corpora is 1. Each comment/post is annotated with sentiment polarity at the comment/post level. This dataset also has the class imbalance problem which is consistent with how sentiments are expressed in the real world. The dataset provided for training and development contains 11,335 and 1,260 sentences for Tamil, 4,851 and 541 sentences for Malayalam. More details about the annotation of the dataset can be found in [13] and [14].

Our shared task aims to encourage research that will reveal how sentiment is expressed in code-mixed scenarios on Dravidian social media text. The participants were provided with development, training and test datasets.

## 2. Task Description

The Dravidian-CodeMix-FIRE 2020 was a message-level polarity classification task. As a part of this shared task, participants were supposed to develop systems that classify a given Youtube comment into one of the five classes: *positive*, *negative*, *neutral*, *mixed emotions*, and *not in the targeted languages (Tamil or Malayalam)*. Our datasets consist switching at three levels - Inter-Sentential, Intra-Sentential and Tag. All comments in the dataset were written in Roman script with either Tamil grammar and English lexicon or English grammar and Tamil lexicon. The following examples from the Tamil dataset illustrate this scripting pattern.

- **Intha padam vantha piragu yellarum Thala ya kondaduvanga.** - *After the movie release, everybody will celebrate the hero.* Tamil words written in Roman script with no English switch.
- **Trailer late ah parthavanga like podunga.** - *Those who watched the trailer late, please like it.* Tag switching with English words.

---

<sup>1</sup>Between 2011- 2015 Source [https://en.wikipedia.org/wiki/Tamil\\_language](https://en.wikipedia.org/wiki/Tamil_language)

<sup>2</sup>Between 2011- 2019 Source <https://en.wikipedia.org/wiki/Malayalam>

<sup>3</sup>This was also called Damili or Tamil-Brahmi script

- **Omg .. use head phones. Enna bgm da saami ..** - *OMG! Use your headphones. Good Lord, What a background score!* Inter-sentential switch
- **I think sivakarthisckku hero getup set aagala.** - *I think the hero role does not suit Sivakarthisck.* Intra-sentential switch between clauses.

The following examples from the Malayalam dataset also show a similar scripting pattern.

- **Orupaadu nalukalku shesham aanu ithupoloru padam eranghunnathu.** - *A movie like this is coming out after a long time.* Malayalam words written in Roman script with no English switch.
- **Malayalam industry ku thriller kshamam illannu kaanichu kodukku anghotu.** - *Show that there is no shortage for thriller movies in Malayalam film industry.* Tag switching with English words.
- **Manju chechiyude athyugran performancenayi kaathirikunnu. The Lady superstar of Malayalam industry.** - *Waiting for the awesome performance of Manju sister. The Lady superstar of Malayalam film industry.* Inter-sentential switch
- **Next movie ready for nammude swantham dhanush.** - *Next movie ready for our dear Dhanush.* Intra-sentential switch between clauses.

The data was annotated for sentiments according to the following schema.

- **Positive state:** There is an explicit or implicit clue in the text suggesting that the speaker is in a positive state, i.e., happy, admiring, relaxed, and forgiving.
- **Negative state:** There is an explicit or implicit clue in the text suggesting that the speaker is in a negative state, i.e., sad, angry, anxious, and violent.
- **Mixed feelings:** There is an explicit or implicit clue in the text suggesting that the speaker is experiencing both positive and negative feeling: Comparing two movies
- **Neutral state:** There is no explicit or implicit indicator of the speaker's emotional state: Examples are asking for like or subscription or questions about the release date or movie dialogue. This state can be considered as a neutral state.
- **Not in intended language:** For Malayalam, if the sentence does not contain Malayalam then it is not Malayalam.

The annotators were provided with Tamil and Malayalam translation of the above to facilitate better understanding. Each sentence was annotated by a minimum of three annotators.

### 3. Methodology

We received a total of 32 submissions for Tamil and 28 for Malayalam. The systems were evaluated based on weighted average F1 scores and a rank list was prepared. Table 1 and Table 2 show the rank lists of Tamil and Malayalam tasks respectively. We briefly describe below the methodologies used by the top three teams.

No.	TeamName	Precision	Recall	F1-Score	Rank
01	SRJ[15]	0.64	0.67	0.65	1
02	DT	0.62	0.68	0.64	2
03	YUN111 [16]	0.63	0.67	0.64	2
04	codemixed_umsnh [17]	0.61	0.68	0.63	3
05	LucasHub [18]	0.61	0.68	0.63	3
06	YNU [19]	0.61	0.67	0.63	3
07	MUCS[20]	0.60	0.66	0.62	4
08	PITS [21]	0.62	0.69	0.62	4
09	datamafia	0.60	0.65	0.62	4
10	gauravarora [22]	0.65	0.69	0.62	4
11	jiaming gao	0.61	0.64	0.62	4
12	Theedhum Nandrum[23]	0.64	0.67	0.62	4
13	HRS-TECHIE Tam [24]	0.59	0.65	0.61	5
14	NITP-AI-NLP [25]	0.59	0.64	0.61	5
15	SSNCSE_NLP[26]	0.60	0.65	0.61	5
16	zyy1510 [27]	0.59	0.66	0.61	5
17	bits2020 [28]	0.62	0.66	0.61	5
18	SSN_NLP_MLRG [29]	0.60	0.68	0.60	6
19	Siva [30]	0.59	0.63	0.60	6
20	IRLab@IITV	0.59	0.61	0.59	7
21	CMSAOne [31]	0.58	0.67	0.58	8
22	CodeMixedNLP_submission	0.56	0.68	0.58	8
23	ComMA	0.58	0.66	0.58	8
24	IRLab@IITBHU Both [32]	0.57	0.61	0.58	8
25	JUNLP[33]	0.59	0.66	0.58	8
26	TADS [34]	0.57	0.67	0.56	9
27	Parameswari_Faith_Nagaraju [35]	0.55	0.66	0.55	10
28	Judith Jeyafreeda [36]	0.57	0.66	0.54	11
29	Thirumurugan R	0.67	0.66	0.54	11
30	DLRG	0.62	0.49	0.53	12
31	NUIG_Shubhanker [37]	0.52	0.52	0.51	13
32	Anbukkarasi [38]	0.33	0.07	0.10	14

**Table 1**

Rank list based on weighted average F1-score along with other evaluation metrics (Precision and Recall) for Tamil track

- SRJ[15]: Authors used XLM-Roberta’s hidden states to extract semantic information. They proposed a new model by extracting the output of top hidden layers in XLM-Roberta and feeding them as inputs to a Convolution Neural Network and finally concatenate them to get better results. They achieved the best result for both Tamil and Malayalam.
- YUN111 [16]: This team used mBERT to represent the code-mixed Dravidian text, which has been feed to the BiLSTM (creates attention weighted vector representation of the vector). In the end, the outputs of BiLSTM and attention layer of mBERT are concatenated for the classification. The system achieved Rank 2 for Tamil and Malayalam.
- codemixed\_umsnh [17]: Authors combined several models that solved the task separately; They then made a final decision through differential evolution and a linear combination of independently computed decision-value of each model. This system achieved 3rd place for Tamil and

No.	TeamName	Precision	Recall	F1-Score	Rank
01	SRJ [15]	0.74	0.75	0.74	1
02	datamafia	0.74	0.74	0.74	1
03	YNU [19]	0.74	0.74	0.74	1
04	YUN111 [16]	0.73	0.73	0.73	2
05	LucasHub [18]	0.73	0.73	0.73	2
06	jiaming gao	0.73	0.73	0.73	2
07	DT	0.72	0.72	0.72	3
08	CIA_NITT [39]	0.71	0.71	0.71	4
09	PITS [21]	0.70	0.71	0.71	4
10	SSNCSE_NLP [26]	0.71	0.71	0.71	4
11	NITP-AI-NLP [25]	0.69	0.69	0.69	5
12	gauravarora [22]	0.69	0.70	0.69	5
13	MUCS [20]	0.68	0.68	0.68	6
14	codemixed_umsnh [17]	0.68	0.69	0.68	6
15	TADS [34]	0.68	0.68	0.67	7
16	CMSAOne [31]	0.66	0.67	0.66	8
17	Siva [30]	0.67	0.67	0.66	8
18	Theedhum Nandrum [23]	0.67	0.66	0.65	9
19	ComMA	0.64	0.66	0.64	10
20	zyy1510 [27]	0.64	0.64	0.64	10
21	IRLab@IITBHU [32]	0.63	0.64	0.63	11
22	CodeMixedNLP_submission	0.59	0.62	0.60	12
23	IRLab@IITV	0.68	0.60	0.60	12
24	SSN_NLP_MLRG [29]	0.61	0.61	0.60	12
25	bits2020 [28]	0.67	0.59	0.60	12
26	Judith Jeyafreeda [36]	0.68	0.62	0.58	13
27	Parameswari_Faith_Nagaraju [35]	0.53	0.51	0.48	14
28	NUIG_Shubhanker [37]	0.48	0.50	0.46	15

**Table 2**

Rank list based on weighted average F1-score along with other evaluation metrics (Precision and Recall) for Malayalam track

6th place for Malayalam.

- LucasHub [18]: This team used a multi-step integration method using M-BERT and XLM-RoBERTa. They ranked 2nd and 3rd for Malayalam and Tamil respectively.
- YNU [19]: The system proposed by the team is based on a pre-trained multi-language model XLM-RoBERTa, and uses the K-folding method to the ensemble that aims to solve the sentiment analysis problem of multilingual code-mixed across language models. This system achieved rank 1 for Malayalam and 3 for Tamil.

## 4. Evaluation

The distribution of the sentiment classes are imbalanced in both the datasets. In the Malayalam-English code-mixed dataset, we have a class imbalance with the majority of comments belonging to positive (2,811) and neutral (1,903) classes. Similarly, the Tamil-English code-mixed dataset has class imbalance with Positive (10,559), Negative (2,037) and Mixed feelings (1,801) being the majority

classes. This imbalance demands to be addressed. Hence, we chose a weighted average F1-score to rank the systems. The weighted average F1-score is calculated by averaging the support-weighted mean per-class F1 scores (i.e. weights on class distribution). This takes into account the varying degrees of importance of each class in the dataset. We used a classification report tool from Scikit learn<sup>4</sup>.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F-Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

## 5. Results and Discussion

Overall, 119 participants registered for this track. 32 teams submitted final results for Tamil and 28 teams submitted results for Malayalam. Table 1 and Table 2 show the rank list of Tamil and Malayalam task respectively. The runs are sorted in decreasing order of the weighted F1-scores. It is noteworthy that most of the participants used pre-trained embedding such as BERT or its variations even though BERT or its variations are not trained on code-mixed text. Since our corpus contained text written in non-native script, the choice runs counter to linguistic intuitions. There were some systems based on BiLSTM and Recurrent Neural Networks (RNNs). A few other submissions adopted linguistically motivated methods to solve the problem. However they did not achieve good results compared to the BERT based models. Out of all the models proposed, the count vectorization model and the BERT-based model produced the best outcomes. Although there were many systems that were below the baseline results, the approaches taken by participants were different and hence we accepted those papers as well in order to encourage diverse research methods to solve the problem. Although weighted scores were considered as the primary metric for our evaluation, it can be noted that class-wise precision, recall, and F1-Score were reported in the most of papers for better understanding of the problem and results.

Some of the participants made interesting observations about the dataset provided by us in and were able to explain the low F1 scores based on that. Although the data was annotated by a minimum of three annotators and an inter-annotator agreement of 0.6 for Tamil and 0.8 for Malayalam in Krippendorff’s alpha was achieved, the dataset contained instances of annotation errors which were pointed out by Krishnamurthy et. al [35] and BalaSundaraRaman et. al [23]. According to Krishnamurthy et al, a few Malayalam sentences belonging to the *positive* class were wrongly annotated as Not-Malayalam. They also pointed out that some sentences were wrongly tagged as Negative while they actually expressed positive sentiments. According to BalaSundaraRaman et. al [23], there were mismatched predictions in Tamil development dataset, where the authors’ algorithm made correct predictions, but the corresponding manual labels given by the annotators was wrong. To ensure high quality annotation we followed the following protocol. Native language speakers from both the genders were chosen for the task. They were given proper guidelines in their native language and English. We ensured that only after understanding the annotation scheme thoroughly each annotator could proceed to assign the sentiment labels. The manual annotation was carried out in three stages. First, each sentence was annotated by two people. In the second step, the data were collected if both of

---

<sup>4</sup>[https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification\\_report.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html)

them agreed. In the case of conflict, a third person annotated the sentence. In the third step, if all the three of them did not agree, then two more annotators annotated the sentences. Despite following this strict protocol, errors have occurred in the gold standard dataset. We will consider all these suggestions for next years shared task.

The best performing run achieved weighted F1-score of 0.65 and 0.74 for Tamil and Malayalam respectively. The top team “SRJ” used XLM-Roberta and CNN to propose new model to extract semantic information. These scores are relatively low compared to the monolingual sentiment analysis results in high-resourced languages such as English. Code mixing is a challenging issue since the text is written in non-native script with no standard spelling which causes ambiguities. The word written in non-native script, in our case it is the Latin script, causes variable lexical representations. Since our corpus contains all three types of code-mixing including code-mixing at the word and morpheme levels word-level, it gives rise to out-of-vocabulary problems. Other challenges related to code-mixing are reduplication of words, variations in word order. The challenges faced by the implementations submitted for the shared task reflect the complexity of code-mixing and class imbalance issues in the real-world setting. Coupled with these challenges is the fact that the shared task was conducted in the under-resourced setting which makes it even more difficult to get high results.

## 6. Conclusion

This paper overviews the first shared task on sentiment analysis in code-mixed Dravidian text from social media that aims at classifying YouTube comments. The hundred and nineteen teams participated in the task, and a total of 32 teams for Tamil and 28 teams Malayalam submitted the results. Systems have been trained on the unbalanced dataset. The methods proposed by participants ranged from traditional machine learning models with features based approaches to using state-of-the-art embedding methods in deep learning models. In future, we plan to extend the task to other Dravidian languages such as Kannada, Telugu, and Tulu. We also plan to include mixed script data to make the system more real time.

## Acknowledgments

This publication is the outcome of the research supported in part by a research grant from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289\_P2 (Insight\_2), co-funded by the European Regional Development Fund as well as by the EU H2020 programme under grant agreements 825182 (Prêt-à-LLOD), and Irish Research Council grant IRCLA/2017/129 (CARDAMOM-Comparative Deep Models of Language for Minority and Historical Languages).

## References

- [1] N. Jose, B. R. Chakravarthi, S. Suryawanshi, E. Sherly, J. P. McCrae, A survey of current datasets for code-switching research, in: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020.
- [2] R. Priyadharshini, B. R. Chakravarthi, M. Vegupatti, J. P. McCrae, Named entity recognition for code-mixed Indian corpus using meta embedding, in: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020.
- [3] K. Bali, J. Sharma, M. Choudhury, Y. Vyas, “I am borrowing ya mixing ?” an analysis of English-Hindi code mixing in Facebook, in: Proceedings of the First Workshop on Computational Ap-



- proaches to Code Switching, Association for Computational Linguistics, Doha, Qatar, 2014, pp. 116–126. URL: <https://www.aclweb.org/anthology/W14-3914>. doi:10.3115/v1/W14-3914.
- [4] T. Dhanabalan, R. Parthasarathi, T. Geetha, Tamil spell checker, in: Sixth Tamil Internet 2003 Conference, Chennai, Tamilnadu, India, 2003.
- [5] B. Premjith, K. Soman, M. A. Kumar, A deep learning approach for malayalam morphological analysis at character level, *Procedia computer science* 132 (2018) 47–54.
- [6] B. R. Chakravarthi, M. Arcan, J. P. McCrae, Improving Wordnets for Under-Resourced Languages Using Machine Translation, in: Proceedings of the 9th Global WordNet Conference, The Global WordNet Conference 2018 Committee, 2018. URL: [http://compling.hss.ntu.edu.sg/events/2018-gwc/pdfs/GWC2018\\_paper\\_16](http://compling.hss.ntu.edu.sg/events/2018-gwc/pdfs/GWC2018_paper_16).
- [7] B. R. Chakravarthi, M. Arcan, J. P. McCrae, WordNet gloss translation for under-resourced languages using multilingual neural machine translation, in: Proceedings of the Second Workshop on Multilingualism at the Intersection of Knowledge Bases and Machine Translation, European Association for Machine Translation, Dublin, Ireland, 2019, pp. 1–7. URL: <https://www.aclweb.org/anthology/W19-7101>.
- [8] M. Anand Kumar, V. Dhanalakshmi, K. Soman, S. Rajendran, A sequence labeling approach to morphological analyzer for Tamil language, *IJCSE) International Journal on Computer Science and Engineering* 2 (2010) 194–195.
- [9] B. R. Chakravarthi, M. Arcan, J. P. McCrae, Comparison of Different Orthographies for Machine Translation of Under-Resourced Dravidian Languages, in: 2nd Conference on Language, Data and Knowledge (LDK 2019), volume 70 of *OpenAccess Series in Informatics (OASICs)*, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2019, pp. 6:1–6:14. URL: <http://drops.dagstuhl.de/opus/volltexte/2019/10370>. doi:10.4230/OASICs.LDK.2019.6.
- [10] B. Krishnamurti, *The Dravidian languages*, Cambridge University Press, 2003.
- [11] B. R. Chakravarthi, N. Rajasekaran, M. Arcan, K. McGuinness, N. E. O’Connor, J. P. McCrae, Bilingual lexicon induction across orthographically-distinct under-resourced Dravidian languages, in: Proceedings of the Seventh Workshop on NLP for Similar Languages, Varieties and Dialects, Barcelona, Spain, 2020.
- [12] B. R. Chakravarthi, Leveraging orthographic information to improve machine translation of under-resourced languages, Ph.D. thesis, NUI Galway, 2020.
- [13] B. R. Chakravarthi, V. Muralidaran, R. Priyadharshini, J. P. McCrae, Corpus creation for sentiment analysis in code-mixed Tamil-English text, in: Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL), European Language Resources association, Marseille, France, 2020, pp. 202–210. URL: <https://www.aclweb.org/anthology/2020.sltu-1.28>.
- [14] B. R. Chakravarthi, N. Jose, S. Suryawanshi, E. Sherly, J. P. McCrae, A sentiment analysis dataset for code-mixed Malayalam-English, in: Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL), European Language Resources association, Marseille, France, 2020, pp. 177–184. URL: <https://www.aclweb.org/anthology/2020.sltu-1.25>.
- [15] R. Sun, X. Zhou, SRJ @ Dravidian-CodeMix-FIRE2020: Automatic Classification and Identification Sentiment in Code-mixed Text, in: FIRE (Working Notes), 2020.
- [16] Y. Zhu, K. Dong, YUN111@Dravidian-CodeMix-FIRE2020: Sentiment Analysis of Dravidian Code Mixed Text, in: FIRE (Working Notes), 2020.
- [17] J. Ortiz-Bejar, J. Ortiz-Bejar, J. Cerda-Jacobo, M. Graff, E. S. Tellez, UMSNH-INFOTEC@Dravidian-CodeMix-FIRE2020: An ensemble approach based on a multiple text rep-



- resentations, in: FIRE (Working Notes), 2020.
- [18] B. Huang, Y. Bai, LucasHub@Dravidian-CodeMix-FIRE2020: Sentiment Analysis on Multilingual Code Mixing Text with M-BERT and XLM-RoBERTa, in: FIRE (Working Notes), 2020.
  - [19] X. Ou, H. Li, YNU@Dravidian-CodeMix-FIRE2020: XLM-RoBERTa for Multi-language Sentiment Analysis, in: FIRE (Working Notes), 2020.
  - [20] F. Balouchzahi, H. L. Shashirekha, MUCS@Dravidian-CodeMix-FIRE2020:SACO-SentimentsAnalysis for CodeMix Text, in: FIRE (Working Notes), 2020.
  - [21] N. Kanwar, M. Agarwal, R. K. Mundotiya, PITS@Dravidian-CodeMix-FIRE2020: Traditional Approach to Noisy Code-Mixed Sentiment Analysis, in: FIRE (Working Notes), 2020.
  - [22] G. Arora, Gauravarora@HASOC-Dravidian-CodeMix-FIRE2020: Pre-training ULMFiT on Synthetically Generated Code-Mixed Data for Hate Speech Detection, in: FIRE (Working Notes), 2020.
  - [23] B. L. S. K. Ravindranath, Theedhum Nandrum@Dravidian-CodeMix-FIRE2020: A Sentiment Polarity Classifier for YouTube Comments with Code-switching between Tamil, Malayalam and English, in: FIRE (Working Notes), 2020.
  - [24] S. Swaminathan, H. K. Ganesan, R. Pandiyarajan, HRS-TECHIE@Dravidian-CodeMix-FIRE2020 Social Media Sentiment Analysis for Dravidian Languages using Machine Learning, Deep Learning and Ensemble Approaches, in: FIRE (Working Notes), 2020.
  - [25] A. Kumar, S. Saumya, J. P. Singh, NITP-AI-NLP@Dravidian-CodeMix-FIRE2020: A Hybrid CNN and Bi-LSTM Network for Sentiment Analysis of Dravidian Code-Mixed Social Media Posts, in: FIRE (Working Notes), 2020.
  - [26] N. N. Appiah Balaji, B. B, B. J, SSNCSE\_NLP@Dravidian-CodeMix-FIRE2020: Sentiment Analysis for Dravidian Languages in Code-Mixed Text, in: FIRE (Working Notes), 2020.
  - [27] Y. Zhu, X. Zhou, Zyy1510@HASOC-Dravidian-CodeMix-FIRE2020: An Ensemble Model for Offensive Language Identification, in: FIRE (Working Notes), 2020.
  - [28] Y. Sharma, A. V. Mandalam, BITS2020@Dravidian-CodeMix-FIRE2020: Sub-Word Level Sentiment Analysis of Dravidian Code Mixed Data, in: FIRE (Working Notes), 2020.
  - [29] A. Kalaivani, D. Thenmozhi, SSN\_NLP\_MLRG@Dravidian-CodeMix-FIRE2020: Sentiment Code-Mixed Text Classification in Tamil and Malayalam using ULMFiT, in: FIRE (Working Notes), 2020.
  - [30] S. Sai, Y. Sharma, Siva@HASOC-Dravidian-CodeMix-FIRE-2020: Multilingual Offensive Speech Detection in Code-mixed and Romanized Text, in: FIRE (Working Notes), 2020.
  - [31] S. Dowlagar, R. Mamidi, CMSAOne@Dravidian-CodeMix-FIRE2020: A Meta Embedding and Transformer model for Code-Mixed Sentiment Analysis on Social Media Text, in: FIRE (Working Notes), 2020.
  - [32] S. Chanda, S. Pal, IRLab@IITBHU@Dravidian-CodeMix-FIRE2020: Sentiment Analysis for Dravidian Languages in Code-Mixed Text, in: FIRE (Working Notes), 2020.
  - [33] S. K. Mahata, D. Das, S. Bandyopadhyay, JUNLP@Dravidian-CodeMix-FIRE2020: Sentiment Classification of Code-Mixed Tweets using Bi-Directional RNN and Language Tags, in: FIRE (Working Notes), 2020.
  - [34] D. Sharma, TADS@Dravidian-CodeMix-FIRE2020: SentimentAnalysisonCodeMixDravidian-Language, in: FIRE (Working Notes), 2020.
  - [35] P. Krishnamurthy, F. Varghese, N. Vuppala, Parameswari\_faith\_nagaraju@Dravidian-CodeMix-FIRE2020: A machine-learning approach using n-grams in sentiment analysis for code-mixed texts:A case study in Tamil and Malayalam, in: FIRE (Working Notes), 2020.
  - [36] J. Jeyafreeda, JudithJeyafreeda@Dravidian-CodeMix-FIRE2020:Sentiment Analysis of YouTube Comments for DravidianLanguages, in: FIRE (Working Notes), 2020.

- [37] S. Banerjee, A. Jaypal, S. Thavareesan, NUIG-Shubhanker@Dravidian-CodeMix-FIRE2020: Sentiment Analysis of Code-Mixed Dravidian text using XLNet, in: FIRE (Working Notes), 2020.
- [38] A. S, V. S, SA\_SVG@Dravidian-CodeMix-FIRE2020: Deep Learning Based Sentiment Analysis in Code-mixed Tamil-English Text, in: FIRE (Working Notes), 2020.
- [39] Y. Prakash Babu, R. Eswari, K. Nimmi, CIA\_NITT@Dravidian-CodeMix-FIRE2020: Malayalam-English Code Mixed Sentiment Analysis Using Sentence BERT And Sentiment Features, in: FIRE (Working Notes), 2020.