GunadarmaXBRIN at SemEval-2023 Task 12: Utilization of SVM and AfriBERTa for Monolingual, Multilingual, and Zero-shot Sentiment Analysis in African Languages

Novitasari Arlim
Gunadarma University
Depok, Indonesia
novitasariarlim@gmail.com

Slamet Riyanto
National Research and Innovation Agency
Jakarta, Indonesia
slam011@brin.go.id

Rodiah Rodiah
Gunadarma University
Depok, Indonesia
rodiah@staff.gunadarma.ac.id

Al Hafiz Akbar Maulana Siagian
National Research and Innovation Agency
Jakarta, Indonesia
alha001@brin.go.id

Abstract

This paper describes our participation in Task 12: AfriSenti-SemEval 2023, i.e., Track 12 of subtask A, Track 16 of subtask B, and Track 18 of subtask C. To deal with these three tracks, we utilize Support Vector Machine (SVM) + One vs Rest, SVM + One vs Rest with SMOTE, and AfriBERTa-large models. In particular, our SVM + One vs Rest with SMOTE model could obtain the highest weighted F1-Score for Tracks 16 and 18 in the evaluation phase, that is, 65.14% and 33.49%, respectively. Meanwhile, our SVM + One vs Rest model could perform better than other models for Track 12 in the evaluation phase.

1 Introduction

Sentiment Analysis (SA) is one of the Natural Language Processing (NLP) tasks. SA is used to recognize and classify the author's opinion or experience of an object, such as a specific topic, product, service, organization, social events, politics, and economic (Nahili et al., 2021; Yang, 2021; Mehmood et al., 2019). The sentiment of the opinion could be classified as positive, neutral, or negative. Implementation of SA is employed in various sectors, such as business, healthcare, and education (Nandwani and Verma, 2021). However, SA research mostly focused on high-resource language, i.e., English and Chinese (Nasim and Ghani, 2020). In other words, low-resource language is still not yet addressed well (Muhammad et al., 2022), such as African language (Muhammad et al., 2023b), which is the focus of task 12.

Task 12 AfriSenti-SemEval[1] (Muhammad et al., 2023b) aims to do sentiment analysis on African datasets collected from Twitter. This task 12 consists of three subtasks, i.e., monolingual sentiment classification (subtask A), multilingual sentiment classification (subtask B), and zero-shot sentiment classification (subtask C). In total, there are 14 datasets provided in the task 12, namely, Hausa, Yoruba, Igbo, Nigerian Pidgin, Amharic, Algerian Arabic, Moroccan Arabic/Darija, Swahili, Kinyarwanda, Twi, Mozambican Portuguese, Xitsonga (Mozambique Dialect), Tigrinya, and Oromo.

In this AfriSenti-SemEval task, we participate in Track 12: Xitsonga (Mozambique Dialect) of subtask A, Track 16: 12 languages in subtask A of subtask B, and Track 18: Zero-Shot on Oromo of subtask C. To deal with each subtask, we utilize three models, namely, Support Vector Machine (SVM) + One vs Rest (Model 1), SVM + One Vs Rest with SMOTE (Model 2), and AfriBERTa-large (Ogueji et al., 2021) (Model 3). The highest-obtained F1-Scores of our participation are 50.56%, 65.14%, and 33.49% for Track 12, Track 16, and Track 18, respectively. The code used in this work is available on GitHub[2].

2 Background

2.1 Dataset Description

The organizer of the AfriSenti-SemEval task provides training, development, and evaluation datasets in subtasks A and B. Note that the training dataset in subtask B is the combination of all training datasets in subtask A. On the other hand, the organizer prepares only development and evaluation datasets in subtask C. Thus, participants can utilize any or all training datasets available in

[1] https://codalab.lisn.upsaclay.fr/competitions/7320

[2] https://github.com/aquemos/GunadarmaXBRIN-at-SemEval-2023-task-12

Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), pages 869–877
July 13-14, 2023 ©2023 Association for Computational Linguistics

subtask A as their training dataset for subtask C. The provided training datasets consist of ID, tweet, and label. Table 1 shows the data distribution of each class for each track in subtask A. Meanwhile, Table 2 exhibits the number of development and evaluation datasets of each subtask that we follow.

2.2 Related Work

Utilization of fine-tuning method on several pre-trained models obtained a good result for SA in African languages. For example, Fourati et al. (2021) used mBERT and mBERT with Convolutional Neural Network (CNN) models to perform SA on Tunizi dataset. They obtained a macro F1-score of 0.558 and 0.580 when using the mBERT model and mBERT with CNN one, respectively. The mBERT model was also employed by Martin et al. (2021) to do SA on the Swahili dataset, in which they obtained an F1-score of 0.91 and 0.81 on negative and positive classes, respectively. Another fine-tuning method is AfriBERTa used in Muhammad et al. (2022) for monolingual, multilingual, and zero-shot SA on Hausa, Igbo, Yoruba, and Nigerian Pigdin datasets. In particular, the work of Muhammad et al. (2022) obtained a weighted F1-score around 16-45% for monolingual SA, weighted F1-score of 78.3% for multilingual SA, and an average F1-score of 49.9% for zero-shot SA. Since the focus of the AfriSenti-SemEval task is similar to that of Muhammad et al. (2022), i.e., monolingual, multilingual, and zero-shot SA, we consider employing the AfriBERTa-large in this work. However, the number of African languages in our work is much more than that of in the Muhammad et al.'s (2022) one, viz., Hausa, Yoruba, Igbo, Nigerian Pidgin, Amharic, Algerian Arabic, Moroccan Arabic/Darija, Swahili, Kinyarwanda, Twi, Mozambican Portuguese, Xitsonga, and Oromo.

SVM is one of the supervised machine learning algorithms mostly used in SA (Taj and Girisha, 2021). For example, Demircan et al. (2021) used SVM with undersampling for SA in an imbalanced Turkey dataset, i.e., monolingual. They obtained F1-scores of 0.8, 0.75, and 0.92 in negative, neutral, and positive classes, respectively. On the other hand, using SVM also worked well in the multilingual imbalanced dataset consisting of German, French, and English (Pustulka-Hunt et al., 2018). In particular, Pustulka-Hunt et al. (2018) obtained an F1-score of 0.913 using 10-fold cross-validation in their work. Based on the good result of Demircan et al. (2021) and Pustulka-Hunt et al. (2018), we are motivated to employ SVM for monolingual and multilingual SA in this work. Since SVM is a binary linear classifier (Khan et al., 2022) and the dataset in this AfriSenti-SemEval task is multiclass, we employ One Vs Rest strategy on SVM. Moreover, training datasets in subtask A are imbalanced. Thus, classifiers could not work well in such unequal distributions (Flores et al., 2018). Therefore, we utilize an oversampling technique, i.e., SMOTE, on the minority class to the training datasets in subtask A.

3 Method

Details of preprocessing, models, and evaluation employed in this work are as follows:

3.1 Dataset Preprocessing

Diacritic plays an important role in the African language because the diacritic's position can affect the meaning of words (Nwankwo, 2021). For example, *Àwon omó fo abó* (The children washed the dishes) and *Àwon omó fó abó* (The children broke the dishes) have different meanings in Yoruba (Lanfrica, 2022). For this reason, we do not remove punctuation marks in our preprocessing step. Specifically, the preprocessing step is deleting emojis, mentions, URLs, hashtags, and extra whitespaces. It also includes converting tweets into lowercase. This preprocessing step is applied to the training, development, and evaluation datasets.

3.2 Models

3.2.1 Baseline

Fine-tuning on pre-trained models is conducted as baseline models by the organizer of the AfriSenti-SemEval task. The pre-trained models used in Muhammad et al. (2023a) are AfriBERTa-large, XLM-R-base, AfroXLMR-base, mDeBERTaV3-base, XLM-T-base, XLM-R-large, and AfroXLMR-large. The baseline models are divided into three: (1) the pre-trained model is trained with training data from a target language for the monolingual baseline, (2) a multilingual dataset of all 12 languages is used as the training dataset for the multilingual baseline, and (3) training dataset from any of the 12 languages or multilingual dataset is used as a training dataset for zero-shot baseline (Muhammad et al., 2023a).

Label	am	dz	ha	ig	kr	ma	pcm	pt	sw	ts	twi	yo	Total	%
Positive	1332	417	4687	3084	899	1758	1808	681	547	384	1644	3542	20783	32.63
Neutral	3104	342	4912	4508	1257	2161	72	1600	1072	136	522	3108	22794	35.80
Negative	1548	892	4573	2600	1146	1664	3241	782	191	284	1315	1872	20108	31.57
Total	5984	1651	14172	10192	3302	5583	5121	3063	1810	804	3481	8522	63685	

Table 1: The data distribution of each class of each track in subtask A. am=Amharic, dz=Algerian Arabic, ha=Hausa, ig=Igbo, kr=Kinyarwanda, ma=Darija/Morrocan Arabic, pcm=Nigerian Pidgin, pt=Mozambique Portuguese, sw=Swahili, ts=Xitsonga, twi=Twi, yo=Yoruba

	Track 12	Track 16	Track 18
Dev	203	13653	396
Eval	254	30211	2096

Table 2: The number of development and evaluation datasets of subtask A (Track 12), subtask B (Track 16), and subtask C (Track 18). Dev = Development, Eval = Evaluation

Language	Source	Size
Algerian Arabic	Leipzig	894.7 MB
Moroccan Arabic	Leipzig	245.2 MB
Akan	Leipzig	500 KB
Portuguese	Leipzig	519.4 MB
Xitsonga	Leipzig and Sadilar	8.5 MB
Oromo	GitHub and Leipzig	1 MB
Kinyarwanda	Leipzig	16.6 MB

Table 3: List of corpora used to train AfriBERTa-large tokenizer

3.2.2 AfriBERTa-large

AfriBERTa-large[3] is one of the pre-trained multilingual models which covers 11 African languages, viz., Oromo, Amharic, Gahuza, Hausa, Igbo, Nigerian Pidgin, Somali, Swahili, Tigrinya, and Yoruba (Ogueji et al., 2021). Thus, we use Hugging Face Transformers to train the AfriBERTa-large tokenizer and fine-tune the AfriBERTa-large model to deal with this AfriSenti-Semeval task.

Since the AfriBERTa-large tokenizer still does not support Algerian Arabic, Moroccan Arabic, Twi, Mozambican Portuguese, Xitsonga, and Oromo languages, we train the AfriBERTa-large tokenizer with a corpus collected from several sources, i.e., Leipzig[4], Sadilar[5], and GitHub[6] (see Table 3). We also train the AfriBERTa-large tokenizer with Kinyarwanda corpus because it just supports a code-mixed language of Kinyarwanda and Kirundi (Gahuza) (Ogueji et al., 2021). Note that we can not find a corpus for Twi and Mozambican Portuguese languages. However, according to University of Cambridge Language Centre Resources; Rutgers, the Akan language belongs to the Twi language. Therefore, we use the Akan corpus to train the Twi language. Meanwhile, Portuguese corpus is used to train Mozambican Portuguese since Portuguese is the national language in Mozambique (UNICEF).

In particular, we fine-tune the preprocessed training dataset of Xitsonga to participate in Track 12: Xitsonga (Mozambique Dialect) of subtask A, which is monolingual SA. On the other hand, we use all preprocessed training data of subtask A to fine-tune a model to deal with Track 16 of subtask B and Track 18 of subtask C, which are multilingual and zero-shot SA, respectively. Note that 90% of the training dataset is employed to train the model, while 10% of the training dataset is used for evaluation and optimization during the training process. The detail of training model hyperparameters is in Appendix A.1.

3.2.3 SVM + One Vs Rest

According to Murphy (2018), SVM can deal with multiclass classification problems, that is, by utilizing the One Vs Rest (One Vs All) approach. Thus, we use SVM[7] with One Vs Rest[8] from Scikit Learn in this work.

To deal with Track 12 of subtask A, we train the SVM + One Vs Rest model with text vectorization of preprocessed Xitsonga training dataset. Meanwhile, SVM + One Vs Rest model is trained with text vectorization of preprocessed all training datasets in subtask A to deal with Track 16 of subtask B and Track 18 of subtask C. Text vectoriza-

[3]https://huggingface.co/castorini/afriberta_large
[4]https://wortschatz.uni-leipzig.de/en//download
[5]https://repo.sadilar.org/handle/20.500.12185/364
[6]https://github.com/asmelashteka/HornMT/blob/main/data/orm.txt
[7]https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html
[8]https://scikit-learn.org/stable/modules/generated/sklearn.multiclass.OneVsRestClassifier.html

tion process uses Term Frequency - Inverse Document Frequency (TF-IDF)[9] from Scikit Learn with default parameters. TF-IDF is used based on the successful work of Ahuja et al. (2019) and Dharma and Saragih (2022) that employing TF-IDF with SVM generated good results for sentiment analysis.

3.2.4 SVM + One Vs Rest with SMOTE

Synthetic Minority Oversampling (SMOTE) is one of the oversampling techniques that synthesizes a new example with the existing one (Mutanov et al., 2021). In this work, we utilize SMOTE from Imbalanced Library[10]. This oversampling is applied to all training datasets vectorized with TF-IDF as in 3.2.3. In particular, the oversampled training dataset of Xitsonga is used to train the SVM + One Vs Rest model to deal with Track 12 of subtask A. On the other hand, SVM + One Vs Rest model is trained with all oversampled training datasets in subtask A for Track 16 of subtask B and Track 18 of subtask C.

3.3 Evaluation

The submission of each subtask is evaluated with weighted precision, weighted recall, weighted F1-Score, macro precision, macro recall, and macro F1-Score. The team ranking is ordered by the weighted F1-Score of each team's last submission.

4 Results

Table 4 shows the obtained results of our submissions in the AfriSenti-SemEval task. Specifically, the results are the obtained weighted F1-Score of our participation in Tracks 12, 16, and 18. The results detail can be found in https://afrisenti-semeval.github.io/results/.

Based on the results in Table 4, we note that the best model in the development phase could not be the best model in the evaluation one. For example, our AfriBERTa-large could outperform our SVM + One Vs Rest (with or without SMOTE) in the development phase of Track 18. However, the AfriBERTa-large could not perform better than the SVM + One Vs Rest (with or without SMOTE) in the evaluation phase of Track 18, even worst. We try analyzing this distinction by observing the confusion matrix[11] of our models using Scikit Learn[12]. We found that the AfriBERT-large could predict positive instances better than SVM + One Vs Rest (with or without SMOTE) in the development phase of Track 18. On the other hand, the SVM + One Vs Rest (with or without SMOTE) could classify negative and neutral instances better than the AfriBERT-large in the development phase of Track 18. Note that the number of positive instances in the development phase is larger than that of negative and neutral ones, while the number of positive instances in the evaluation phase is smaller than that of negative and neutral ones. This circumstance might be the reason that the AfriBERTa-large could not perform well in the evaluation phase as in the development one. Moreover, other different characteristics between the development dataset and the evaluation one might affect this inconsistency performance. However, such a presumption should be investigated further in the future.

Table 4 also shows that our AfriBERTa-large could not perform as well as the baseline's AfriBERTa-large in the evaluation phase, even worse. The underperforming of our AfriBERTa-large might be due to the use of additional corpora and the choice of fine-tuning hyperparameters when training the AfriBERTa-large. Nevertheless, it would be better to analyze further this conjecture in the next work.

Overall, our SVM + One Vs Rest (with or without SMOTE) could perform better than our AfriBERTa-large in the evaluation phase of tracks 12, 16, and 18. Sadly, our best-obtained results are still unsatisfactory, even far below the results obtained by the highest-ranking team in tracks 12, 16, and 18. This shortcoming might be due to the dataset characteristic. For example, we found that several Hausa tweets with positive sentiments in the evaluation dataset of Track 16 were misclassified by our models as negative ones. To analyze this misclassification, we extracted the positive and negative words in the tweets using the lexicon sentiment provided by AfriSenti[13]. The extraction result shows that several positive tweets are containing negative words, but positive words do not exist (see Table 5). Moreover, code-mixed between Hausa

[9] https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

[10] https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html

[11] The confusion matrix is available in Appendix A.2

[12] https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html

[13] https://github.com/afrisenti-semeval/afrisent-semeval-2023/tree/main/sentiment_lexicon

	Model	Track 12	Track 16	Track 18
Development	AfriBERTa-large	51.27%	57.62%	**30.08%**
	SVM + One Vs Rest	55.40%	**71.57%**	28.41%
	SVM + One Vs Rest with SMOTE	**59.42%**	71.29%	28.51%
Evaluation	AfriBERTa-large	49.84%	53.38%	26.67%
	SVM + One Vs Rest	**50.56%**	65.13%	32.80%
	SVM + One Vs Rest with SMOTE	50.11%	**65.14%**	**33.49%**
Baseline of evaluation (Muhammad et al., 2023a)	AfriBERTa-large	51.60%	64.70%	N/A
	XLM-R-base	47.40%	64.30%	N/A
	AfroXLMR-base	45.90%	68.40%	N/A
	mDeBERTaV3-base	47.40%	66.10%	N/A
	XLM-T-base	53.80%	65.90%	N/A
	XLM-R-large	43.70%	66.90%	N/A
	AfroXLMR-large	47.30%	71.20%	42.00%

Table 4: The obtained results of our submissions in Tracks 12, 16, and 18 of the AfriSenti-SemEval task. Results in bold indicate the obtained highest weighted F1-Score for each track of our submissions in the Development and Evaluation phases. N/A means the score is not available

No	Positive Tweets	Positive Words	Negative Words
1	make she carry her wahala dey go	-	wahala
2	mehn the cars eeh fada lawd abeg nau me sef wan get car and house for assokoro	-	fada
3	ride on hajiya kin chi uwarsu walahi ba kare mai haushi	-	kare, haushi
4	zamu kara cin banza kenan hala madrid	-	banza
5	wetin den use tey highlight wey con bright like stadium light bayi mama say na mel morgue	-	bayi

Table 5: Extraction of Positive and Negative Words on Hausa Positive Tweets

Words	Hausa Sentiment	Yoruba Sentimen
koya	positive	negative
anfani	positive	positive
gafara	positive	positive
wahala	negative	negative
soke	negative	positive

Table 6: Words that have the same meaning and are written in a similar way in Hausa and Yoruba languages

and English appears in the tweets. Thus, such circumstances might affect the performance of our models. Nonetheless, further investigation should be performed on positive and negative phrases to analyze the classification error.

However, our obtained results might be comparable to the baseline ones since we believe the difference is not significant. This condition might indicate that this task is complicated and challenging. Although each language in the multilingual dataset has different language families, a particular language with particular language families might have a similarity to another one from different language families. For example, the Hausa language is classified as an Afro-Asiatic language family, while the Yoruba language is classified as a Niger-Kongo language family. Although these two languages have different language families, we found in the dataset that 177 words from the two languages are similar. In fact, these similar words could have different sentiments (see Table 6).

5 Conclusion

We utilized SVM + One Vs Rest, SVM + One Vs Rest with SMOTE, and AfriBERTa-large to deal with Tracks 12, 16, and 18 of the AfriSenti-SemEval task. Our highest-obtained weighted F1-Scores were 50.56%, 65.14%, and 33.49% for Tracks 12, 16, and 18, respectively. Although our obtained results indicated that our models could not perform as well as or better than the baseline models, the SVM + One Vs Rest (with or without

SMOTE) results could be compared with those of the baseline ones. For further work, it will be better to explore hyperparameter tuning on the pre-trained model to improve the model performance. In addition, there are several tweets written in Latin script. Thus, it may be beneficial to translate Arabic into Latin and vice versa when dealing with this task in the future.

References

Ravinder Ahuja, Aakarsha Chug, Shruti Kohli, Shaurya Gupta, and Pratyush Ahuja. 2019. The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, 152:341–348. International Conference on Pervasive Computing Advances and Applications- PerCAA 2019.

Murat Demircan, Adem Seller, Fatih Abut, and Mehmet Fatih Akay. 2021. Developing turkish sentiment analysis models using machine learning and e-commerce data. *International Journal of Cognitive Computing in Engineering*, 2:202–207.

Arie Satia Dharma and Yosua Giat Raja Saragih. 2022. Comparison of feature extraction methods on sentiment analysis in hotel reviews. *Sinkron : jurnal dan penelitian teknik informatika*, 7(4):2349–2354.

Andrew Christian Flores, Rogelyn I. Icoy, Christine F. Peña, and Ken D. Gorro. 2018. An evaluation of svm and naive bayes with smote on sentiment analysis data set. In *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*, pages 1–4.

Chayma Fourati, Hatem Haddad, Abir Messaoudi, Moez BenHajhmida, Aymen Ben Elhaj Mabrouk, and Malek Naski. 2021. Introducing a large Tunisian Arabizi dialectal dataset for sentiment analysis. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 226–230, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Ihsan Ullah Khan, Aurangzeb Khan, Wahab Khan, Mazliham Mohd Su'ud, Muhammad Mansoor Alam, Fazli Subhan, and Muhammad Zubair Asghar. 2022. A review of urdu sentiment analysis with multilingual perspective: A case of urdu and roman urdu language. *Computers*, 11(1):3.

Lanfrica. 2022. Lanfrica talks 7 | sentiment analysis in low-resource african languages by shamsuddeen h. muhammad.

Gati L Martin, Medard E Mswahili, and Young-Seob Jeong. 2021. Sentiment classification in swahili language using multilingual bert. *arXiv preprint arXiv:2104.09006*.

Khawar Mehmood, Daryl Essam, Kamran Shafi, and Muhammad Kamran Malik. 2019. Sentiment analysis for a resource poor language—roman urdu. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 19(1).

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Abinew Ali Ayele, Nedjma Ousidhoum, David Ifeoluwa Adelani, Seid Muhie Yimam, Ibrahim Sa'id Ahmad, Meriem Beloucif, Saif M. Mohammad, Sebastian Ruder, Oumaima Hourrane, Pavel Brazdil, Felermino Dário Mário António Ali, Davis David, Salomey Osei, Bello Shehu Bello, Falalu Ibrahim, Tajuddeen Gwadabe, Samuel Rutunda, Tadesse Belay, Wendimu Baye Messelle, Hailu Beshada Balcha, Sisay Adugna Chala, Hagos Tesfahun Gebremichael, Bernard Opoku, and Steven Arthur. 2023a. AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages.

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Seid Muhie Yimam, David Ifeoluwa Adelani, Ibrahim Sa'id Ahmad, Nedjma Ousidhoum, Abinew Ali Ayele, Saif M. Mohammad, Meriem Beloucif, and Sebastian Ruder. 2023b. SemEval-2023 Task 12: Sentiment Analysis for African Languages (AfriSenti-SemEval). In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. Association for Computational Linguistics.

Shamsuddeen Hassan Muhammad, David Ifeoluwa Adelani, Sebastian Ruder, Ibrahim Sa'id Ahmad, Idris Abdulmumin, Bello Shehu Bello, Monojit Choudhury, Chris Chinenye Emezue, Saheed Salahudeen Abdullahi, Anuoluwapo Aremu, Alípio Jorge, and Pavel Brazdil. 2022. NaijaSenti: A Nigerian Twitter sentiment corpus for multilingual sentiment analysis. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 590–602, Marseille, France. European Language Resources Association.

Kevin P Murphy. 2018. *Machine learning: A probabilistic perspective (adaptive computation and machine learning series)*. The MIT Press London, UK.

Galimkair Mutanov, Vladislav Karyukin, and Zhanl Mamykova. 2021. Multi-class sentiment analysis of social media data with machine learning algorithms. *Comput. Mater. Contin*, 69:913–930.

Wedjdane Nahili, Khaled Rezeg, and Okba Kazar. 2021. Sentiment analysis on product reviews data using supervised learning: A comprehensive review of recent techniques. In *Proceedings of the 10th International Conference on Information Systems and Technologies*, ICIST '20, New York, NY, USA. Association for Computing Machinery.

Pansy Nandwani and Rupali Verma. 2021. A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1):81.

Zarmeen Nasim and Sayeed Ghani. 2020. Sentiment analysis on urdu tweets using markov chains. *SN Computer Science*, 1(5):269.

John Justice Nwankwo. 2021. Accentuation: A key factor of native languages in african philosophy. *International Journal of Philosophy*, 9(3):178–181.

Kelechi Ogueji, Yuxin Zhu, and Jimmy Lin. 2021. Small data? no problem! exploring the viability of pretrained multilingual language models for low-resourced languages. In *Proceedings of the 1st Workshop on Multilingual Representation Learning*, pages 116–126, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ela Pustulka-Hunt, Thomas Hanne, Eliane Blumer, and Manuel Frieder. 2018. Multilingual sentiment analysis for a swiss gig. In *2018 6th International Symposium on Computational and Business Intelligence (ISCBI)*, pages 94–98. IEEE.

Rutgers. Akan (twi) at rutgers.

MB Nasreen Taj and GS Girisha. 2021. Insights of strength and weakness of evolving methodologies of sentiment analysis. *Global Transitions Proceedings*, 2(2):157–162.

UNICEF. The impact of language policy and practice on children's learning: Evidence from eastern and southern africa.

University of Cambridge Language Centre Resources. University of cambridge language centre resources - twi (akan, asante).

Zhao Yang. 2021. Sentiment analysis of movie reviews based on machine learning. In *2020 2nd International Workshop on Artificial Intelligence and Education*, WAIE 2020, page 1–4, New York, NY, USA. Association for Computing Machinery.

A Appendix

A.1 Training Hyperparameters

Parameter	**Track 12**	**Tracks 16 and 18**
num_train_epochs	10	1
per_device_train_batch_size	64	32
per_device_eval_batch_size	64	32
eval_steps	10	10
evaluation_strategy	steps	steps
load_best_model_at_end	True	True

Table 7: Hyperparameters to train AfriBERTa-large model for Tracks 12, 16, and 18

A.2 Confusion Matrix

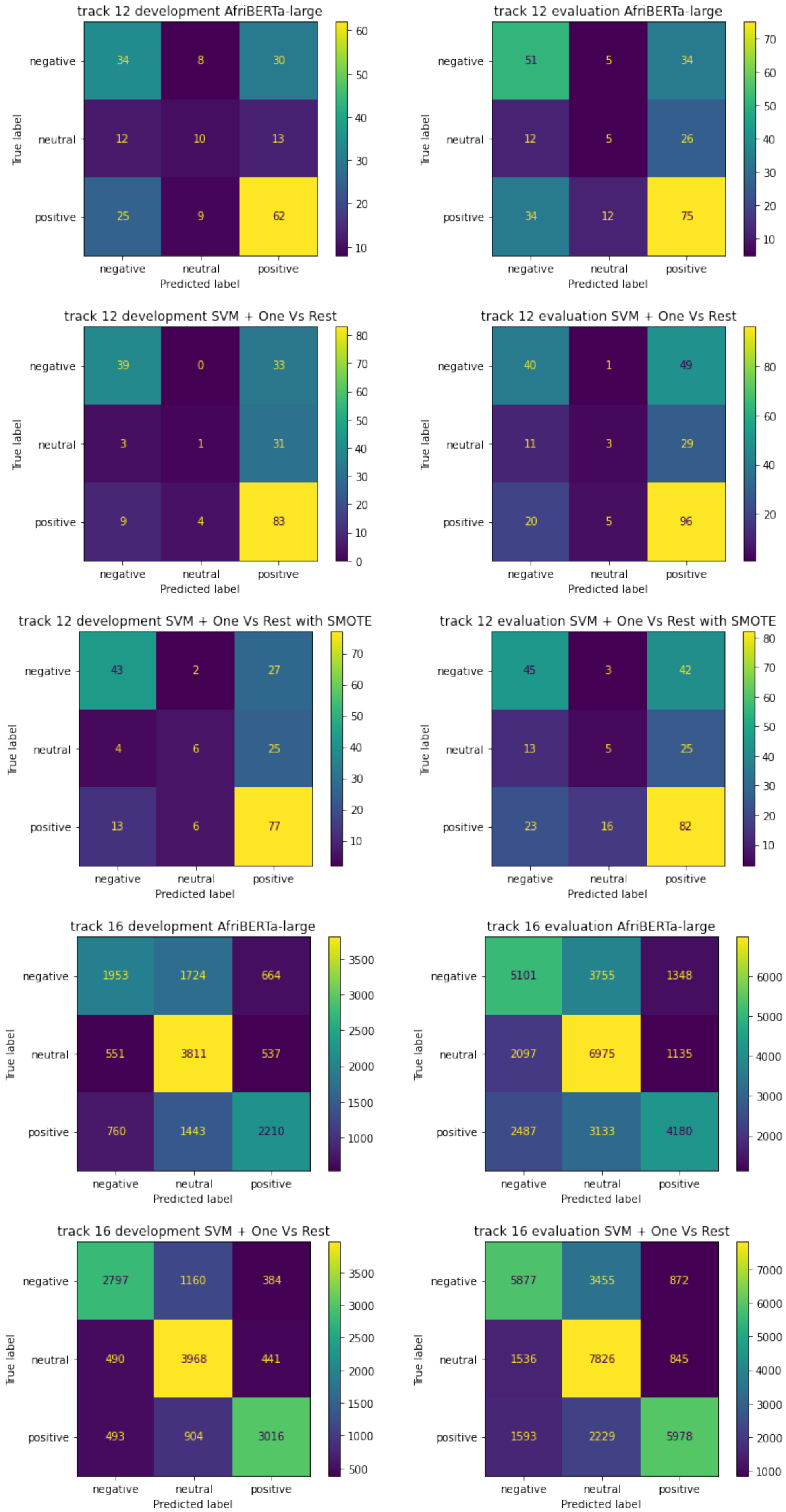

Table 8: Confusion matrix of our models in each track

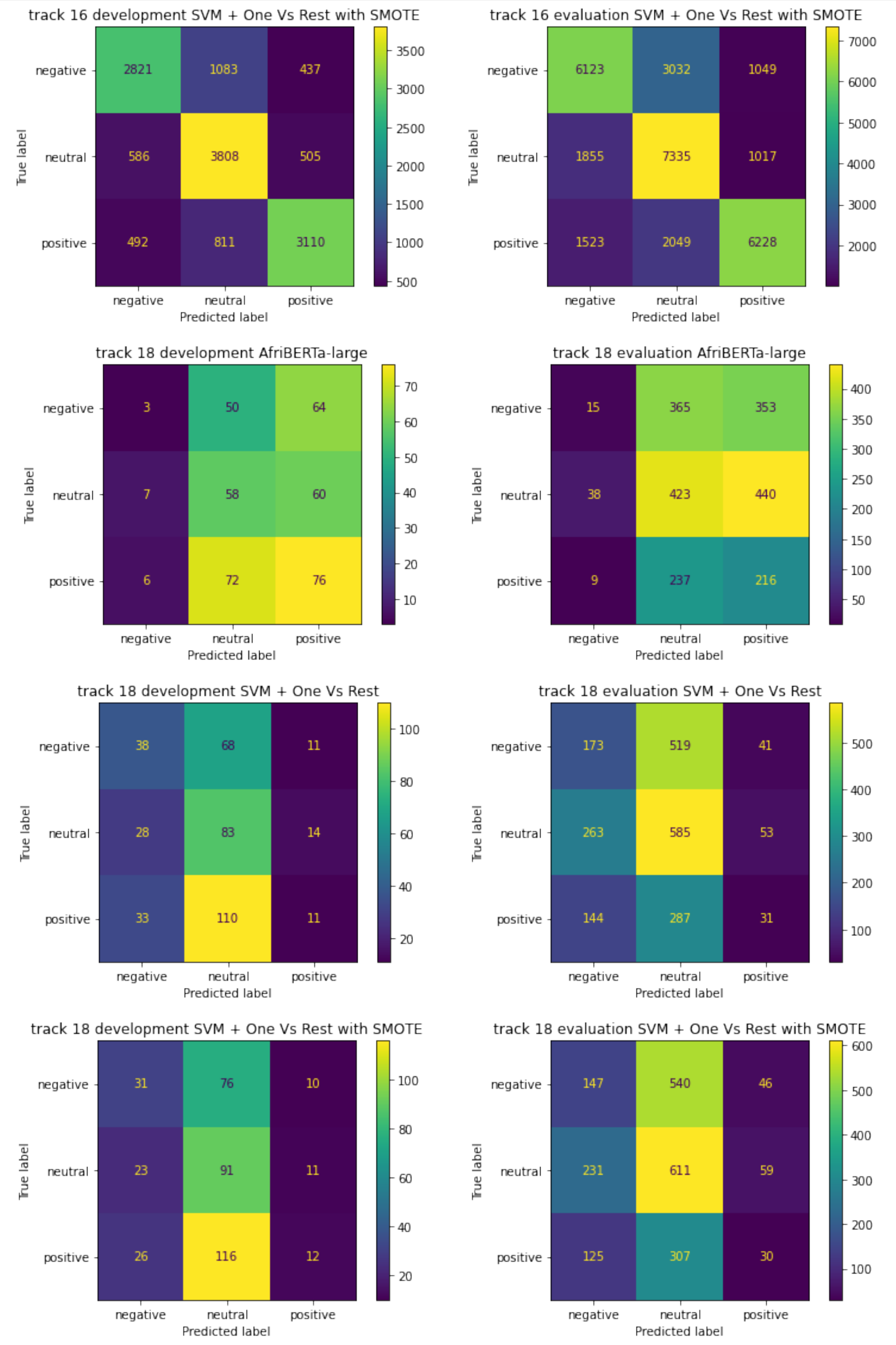

Table 9: Continued: confusion matrix of our models in each track