MIPRO 2021, September 27 - October 01, 2021, Opatija, Croatia

Machine Learning Approaches to Personality Classification on Imbalanced MBTI Datasets

N. Čerkez[*], V. Vareškić [**]
*College of Information Technologies (VSITE), Zagreb, Croatia
** College of Information Technologies (VSITE), Zagreb, Croatia
ninoslav.cerkez@vsite.hr
valentin.vareskic@vsite.hr

***Abstract* - The MBTI (Myers-Briggs Type Indicator) is a widely known approach to personality classification. Datasets for the machine learning approach to personality classification using MBTI are highly imbalanced. Handling imbalanced data sets is a significant open problem with a considerable impact on machine learning methods. This paper presents the results of applying different techniques and suggests their best in mitigating the challenge of imbalanced MBTI datasets. Even though most techniques could be used and implemented to some other problems and areas, like images and sound processing, natural language processing has enough challenges to focus on natural language processing and the specific issue of the MBTI datasets.**

Keywords – machine learning, classification, imbalanced data sets, natural language processing, MBTI

I. INTRODUCTION

Deep learning is the most popular and used technique of machine learning nowadays. It uses multiple processing layers to compose a computational model representing multiple abstraction levels for processed data [1]. Natural Language Processing (NLP) is one of the most prominent fields for implementing machine learning techniques. In the last years, there is a massive demand for improvements in NLP since there is a comprehensive implementation of NLP techniques in industry and everyday life where there is a need to understand and manipulate natural language or speech. One of the tasks in NLP is text classification [2].

A significant open problem in machine learning classification is handling imbalanced data sets. Imbalanced datasets have a considerable impact on the results of machine learning classification since algorithms for machine learning mostly have the presumption that datasets are balanced. However, most datasets are more or less imbalanced. The most challenging problems are naturally imbalanced datasets, like the MBTI dataset, and this paper gives an overview of the present situation and the prospect of future research and implementations focusing on the MBTI dataset.

To demonstrate imbalanced datasets in text classification and different approaches to this problem, we implemented text author Myers-Brigs Type Indicator (MBTI), a personality type classification. We used the PersonalityCafe MBTI dataset[1].

We implement various strategies to deal with imbalanced datasets that can improve multiclass classification of highly imbalanced MBTI datasets by balancing class weights, as suggested in [3] [4].

Data augmentation (DA) relates to strategies that can help deal with imbalanced datasets since they provide increased diversity of training examples. Standard approaches in DA include a combination of oversampling and undersampling techniques. From the perspective of the MBTI dataset and multiclass classification tasks, the most promising approach is Synthetic Minority Oversampling Technique (SMOTE) [5] [6]. As an alternative approach for oversampling, it is possible to use Easy Data Augmentation Technique (EDA) [7]. However, since the MBTI dataset we use has minimal minority classes, undersampling approaches could be an option for larger datasets.

Semantical aspects of errors is a perspective that we would like to take regarding actual and predicted classes, keeping in mind that these classes have string representations and compound structure. So, we analyzed the similarity of MBTI classes, considering that the higher similarity between two classes imposes lower separability of these classes. For that purpose, we used Hamming distance metrics. We wanted to explore if the similarity of classes correlates to Hamming distance between classes. For example, INTJ and ENTJ classes have differences only with the first label, and we explored how their Hamming distance correlates to their semantic textual similarity.

The problem of imbalanced datasets also impacts other areas, like machine learning for processing images [8] [9] or audio processing [10]. This paper emphasizes imbalanced datasets in natural language processing, having the MBTI dataset and personality classification based on the MBTI dataset as a focus.

The paper structure is as follows: Section II describes text processing referring to the chosen dataset; Section III presents experimental setup emphasizing techniques for handling imbalanced multiclass classification on MBTI dataset; Section IV gives the results and discussion, and Section V concludes this paper.

[1] https://www.kaggle.com/datasnaek/mbti-type

2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO) | 978-953-233-101-1/21/$31.00 ©2021IEEE | DOI: 10.23919/MIPRO52101.2021.9596742

II. TEXT PREPROCESSING

We processed the dataset through the standard steps in NLP. We used the nltk package to clean the dataset in machine learning methods and the spacy[2] package PyTorch environment for the neural network approach since it has excellent support for tokenization.

A. Dataset and data background

We used the Kaggle MBTI dataset for this research, consisting of 8.675 rows of posts from the PersonalityCafe forum with labels of 16 MBTI personality types for their authors. We used this dataset because it has exceptionally imbalanced characteristics and offers clear-defined classes for multiclass text classification. The MBTI gives a structure for a psychological classification regarding principal psychological functions, sensation, intuition, feeling, and thinking. Based on Jung's description of personality, MBTI as s psychological instrument includes measures for attitude and functions. Attitude can have values Extrovert (E) and Introvert (I); it measures what an individual prefers: inner or outer world, making the first dichotomy in each personality type. MBTI measures function across the following three dimensions. The second dichotomy can have values Sensation (S) or Intuition (N) and measures information processing; through the five senses or pattern' impressions. The third dimension gives preferences of evaluation others through emotions or principles and facts. It can have values Thinking (T) or Feeling (F). Finally, the fourth dimension measures how individuals prefer to organize their lives as ordered or flexible. It can have values Judgment (J) or Perceiving (P). That gives 16 possible combinations of MBTI personality types [11] [12].

This dataset has a significant imbalance between its classes. For example, only the classes "INFP," "INFJ," "INTP," and "INTJ" have appearance between 12% and 21% in the dataset, the classes "ESFJ," "ESTJ" less than 5%, and all of the other classes between 5% and 8% of corresponding examples in the dataset. Figure 1 shows a bar chart presentation of 16 classes in this dataset.

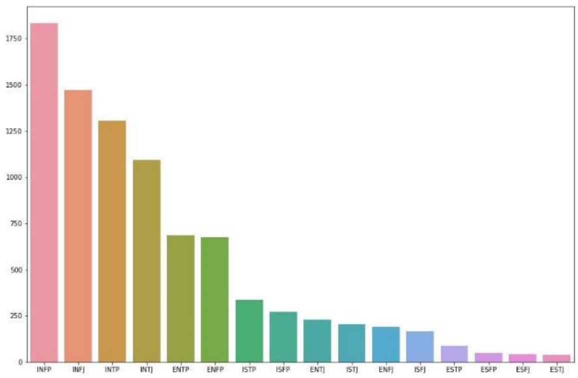

Figure 1. MBTI dataset

For this research, it is helpful to compare the distribution of MBTI classes in the dataset and among the population. Table I shows the distributions, and we can see that MBIT has a naturally imbalanced class distribution [13]. Interestingly, distribution in the dataset and estimated US population has a significant distinction for some classes, and it can be the area of research.

Table I MBTI class distribution

Class	% - dataset	% - Estimated US population
INFP	21.11	4.4% (4-5%)
INFJ	16.95	1.5% (1-3%)
INTP	15.03	3.3% (3-5%)
INTJ	12.58	2.1% (2-4%)
ENTP	7.90	3.2% (2-5%)
ENFP	7.78	8.1% (6-8%)
ISTP	3.89	5.4% (4-6%)
ISFP	3.12	8.8% (5-9%)
ENTJ	2.66	1.8% (2-5%)
ISTJ	2.36	11.6% (11-14%)
ENFJ	2.19	2.5% (2-5%)
ISFJ	1.91	13.8% (9-14%)
ESTP	1.03	4.3% (4-5%)
ESFP	0.55	8.5% (4-9%)
ESFJ	0.48	12.3% (9-13%)
ESTJ	0.45	8.7% (8-12%)

B. Word Embeddings

In the beginning, we prepared text normalization by removing punctuation, special characters, links, stopwords, and numbers. In the next step, we converted uppercase characters to lowercase and removed one-letter words. Next, we used the spacy for lemmatization. Finally, we converted the data into word embeddings using 2 million pre-trained fasttext [3]vectors for some training.

III. EXPERIMENTAL SETUP FOR IMBALANCED DATASETS

This paper aims to give an overview of different approaches to handling imbalanced datasets in natural language processing. Furthermore, we provide a perspective of the present situation and research in this field through experimental results.

Regarding neural network setup, we used a Bidirectional LSTM network [14] [15] with the following parameters: 25 epochs, two layers, fixed input length, the value of learning rate 0.001, the dimension of a hidden layer of 256, the dimension of an embedding of 200, dropout of 0.2 [16], 30% of the dataset for validation phase, maximum vocabulary size of 25.000, batch size of 128, and ADAM as an optimizer [17]. In neural network training, we used Cross-entropy loss as a loss function.

Regarding classic machine learning methods, we used the package sklearn [4] and its support for the following algorithms: Support Vector Classifier from Support Vector Machine, Decision Tree Classifier, Logistic Regression, Ridge Classifier, Multinomial Naïve Bayes Classifier, Random Forest Classifier, XG Boost Classifier, Bagging Classifier, and Ensemble Methods. We trained models on CUDA 10.1 GPUs. Regarding hyperparameters, we used findings in research [18] that suggests parameters we used

[2] https://spacy.io/
[3] https://fasttext.cc/docs/en/english-vectors.html
[4] https://scikit-learn.org/stable/

in our neural networks training. However, our focus was not to achieve the best classification results but to discover the impact of different methods for dealing with imbalanced datasets. Also, we are aware that some strategies we should fine-tune were not in our focus because of the main focus on techniques for dealing with imbalanced datasets.

IV. RESULTS

Table II shows the results of LSTM training without additional techniques for imbalanced datasets. The highly imbalanced dataset primarily directs the results of classification to majority classes. The model did not learn how to classify minority class examples.

Table II LSTM training

Class	Precision	Recall	F1-score	Support
INFP	0.19	0.38	0.25	212
INFJ	0.18	0.62	0.27	171
INTP	0.00	0.00	0.00	159
INTJ	0.00	0.00	0.00	124
ENTP	0.00	0.00	0.00	88
ENFP	0.00	0.00	0.00	86
ISTP	0.00	0.00	0.00	43
ISFP	0.00	0.00	0.00	26
ENTJ	0.00	0.00	0.00	23
ISTJ	0.00	0.00	0.00	20
ENFJ	0.00	0.00	0.00	16
ISFJ	0.00	0.00	0.00	24
ESTP	0.00	0.00	0.00	17
ESFP	0.00	0.00	0.00	3
ESFJ	0.00	0.00	0.00	10
ESTJ	0.00	0.00	0.00	3
accuracy				0.18
macro avg	0.02	0.06	0.03	1025
weighted avg	0.07	0.18	0.10	1025

Synthetic Minority Oversampling Technique (SMOTE) can mitigate the imbalance in datasets [6]. Figure 2 shows training and validation loss for LSTM training with SMOTE. Interestingly, training loss increased from 2.309 to 2.494, validation loss from 2.324 to 2.503, validation accuracy decreased from 0.209 to 0.164, and rmse increased from 3.868 to 5.431. We can explain these results like the fact that the network overfits after 15 epochs.

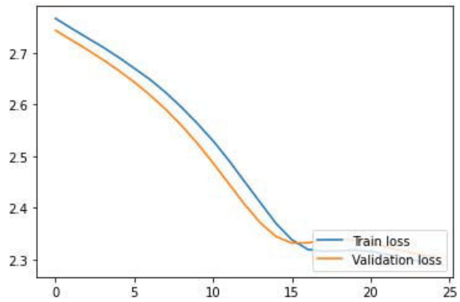

Figure 2. The loss for LSTM with SMOTE

Table III shows the results of better classification for the majority and minority classes. However, training and validation loss are high, and there is much space for improvement.

Table III LSTM with SMOTE

Class	Precision	Recall	F1-score	Support
INFP	0.22	0.85	0.35	241
INFJ	0.16	0.11	0.13	217
INTP	0.00	0.00	0.00	228
INTJ	0.00	0.00	0.00	208
ENTP	0.00	0.00	0.00	207
ENFP	1.00	0.00	0.01	215
ISTP	0.00	0.00	0.00	228
ISFP	0.11	0.12	0.11	214
ENTJ	0.02	0.01	0.01	219
ISTJ	0.10	0.07	0.08	214
ENFJ	0.08	0.03	0.05	220
ISFJ	0.10	0.19	0.13	198
ESTP	0.15	0.14	0.14	213
ESFP	0.18	0.18	0.18	200
ESFJ	0.21	0.50	0.29	230
ESTJ	0.16	0.32	0.21	214
accuracy				0.1
macro avg	0.15	0.16	0.11	3466
weighted avg	0.15	0.16	0.10	3466

Borderline-SMOTE is an approach that focuses only on the examples of the minority class on the border and oversamples only these [19]. Table IV gives the results of LSTM training with this method, which are slightly better. For example, F1-score had the highest value of 0.44 and validation accuracy was 0.208. The graph of the loss for this method gives Figure 3. Again, we can notice that the network has more minor effects of overfitting.

Table IV LSTM with Borderline - SMOTE

Class	Precision	Recall	F1-score	Support
INFP	0.23	0.84	0.36	229
INFJ	0.16	0.12	0.14	203
INTP	0.25	0.01	0.02	209
INTJ	0.00	0.00	0.00	229
ENTP	0.00	0.00	0.00	216
ENFP	0.08	0.01	0.02	220
ISTP	0.12	0.06	0.08	202
ISFP	0.11	0.12	0.11	217
ENTJ	0.16	0.21	0.18	217
ISTJ	0.16	0.14	0.15	217
ENFJ	0.18	0.21	0.20	225
ISFJ	0.24	0.19	0.22	217
ESTP	0.25	0.44	0.32	218
ESFP	0.35	0.61	0.44	214
ESFJ	0.00	0.00	0.00	6
ESTJ	0.00	0.00	0.00	4
accuracy				0.22
macro avg	0.15	0.19	0.14	3043
weighted avg	0.17	0.22	0.16	3043

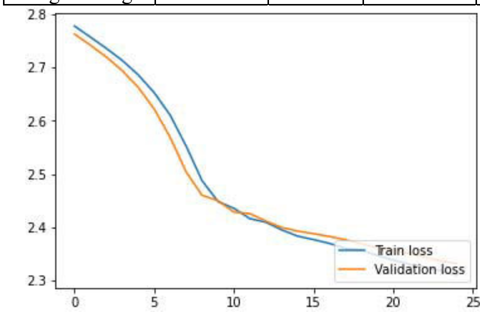

Figure 3. The loss for Borderline- SMOTE

Random oversampling (ROS) and random under-sampling (RUS) are simple data-level methods for addressing the class imbalance where ROS consistently outperforms RUS and baseline methods [20]. Our research also shows better results for the ROS approach.

Table V shows LSTM training results with the ROS approach, with the highest value of F1-score of 0.74 and validation accuracy of 0.35.for the ROS method and just 0.052 for the RUS method.

Table V LSTM with ROS

Class	Precision	Recall	F1-score	Support
INFP	0.05	0.01	0.02	213
INFJ	0.19	0.02	0.04	222
INTP	0.07	0.01	0.02	206
INTJ	0.12	0.03	0.05	218
ENTP	0.12	0.06	0.08	220
ENFP	0.15	0.05	0.08	202
ISTP	0.18	0.22	0.20	228
ISFP	0.20	0.29	0.24	202
ENTJ	0.26	0.27	0.27	212
ISTJ	0.32	0.44	0.37	211
ENFJ	0.31	0.30	0.31	224
ISFJ	0.25	0.32	0.28	218
ESTP	0.36	0.63	0.46	219
ESFP	0.47	0.89	0.61	254
ESFJ	0.69	1.00	0.82	196
ESTJ	0.60	0.95	0.74	221
accuracy				0.35
macro avg	0.27	0.35	0.29	3466
weighted avg	0.27	0.35	0.29	3466

Support vector classifier (SVC) gave much better results with balanced weights parameter than without it. Table VI shows the work for this method. These results were slightly better than the results with the Decision Tree Classifier, where the balanced weights parameter did not improve outcomes.

With the Logistic Regression, we got results in Table VII. We tried Logistic regression with the default parameter and balanced class weights, but the parameter significantly ruined the results. With the Ridge Classifier (with and without balanced weights), Multinomial Naïve Bayes Classifier, Random Forest Classifier (with and without balanced weights), Bagging Classifier, and XG Boost, we got weak results and low scores with the average results in classifying majority classes and poor outcomes in classifying minority classes. With calibration methods with SVC and Random Forrest Classifier, we got just average results.

Table VIII gives an overview of all methods we used in our research and the results. The results we obtained for multiclass text classification from the experiments conducted, in terms of precision, recall, F score, support, and kappa, are presented (we have bolded the best three results per column). The results show that the ROS method outperformed other methods and that SMOTE and Borderline-SMOTE give the second and the third-best result. These results show that the most significant improvement of the baseline model is implementing these three methods.

Table VI SCV with balanced weights

Class	Precision	Recall	F1-score	Support
INFP	0.22	0.13	0.16	222
INFJ	0.18	0.14	0.16	186
INTP	0.15	0.14	0.14	152
INTJ	0.12	0.15	0.13	110
ENTP	0.10	0.15	0.12	85
ENFP	0.06	0.11	0.07	76
ISTP	0.06	0.11	0.07	38
ISFP	0.06	0.08	0.07	37
ENTJ	0.00	0.00	0.00	23
ISTJ	0.00	0.00	0.00	24
ENFJ	0.00	0.00	0.00	20
ISFJ	0.00	0.00	0.00	24
ESTP	1.00	0.10	0.18	10
ESFP	0.00	0.00	0.00	6
ESFJ	0.00	0.00	0.00	8
ESTJ	0.00	0.00	0.00	4
accuracy				0.1
macro avg	0.12	0.07	0.07	1025
weighted avg	0.14	0.12	0.12	1025

Table VII Logistic regression

Class	Precision	Recall	F1-score	Support
INFP	0.22	0.53	0.31	222
INFJ	0.17	0.22	0.19	186
INTP	0.14	0.12	0.13	152
INTJ	0.09	0.06	0.07	110
ENTP	0.17	0.02	0.04	85
ENFP	0.08	0.01	0.02	76
ISTP	0.00	0.00	0.00	38
ISFP	0.00	0.00	0.00	37
ENTJ	0.00	0.00	0.00	23
ISTJ	0.00	0.00	0.00	24
ENFJ	0.14	0.05	0.07	20
ISFJ	0.00	0.00	0.00	24
ESTP	1.00	0.10	0.18	10
ESFP	0.00	0.00	0.00	6
ESFJ	0.00	0.00	0.00	8
ESTJ	0.00	0.00	0.00	4
accuracy				0.18
macro avg	0.13	0.07	0.06	1025
weighted avg	0.14	0.18	0.14	1025

However, the results show better results having in mind constraints with fixed hyperparameters in the experiment setup. Therefore, for better results, we should also experiment with different values for hyperparameters.

In the second part of our research, we used a pre-trained Fasttext model using different regularization approaches in the LSTM neural network. We are aware that regularization is not an option but a must-have, and we wanted to combine different approaches. As a regularisation method, as we mentioned in the experimental setup, we always used dropout, and then we combined L1 and L2 regularization as additional methods.

Table VIII Multiclass Classification Results

Method	Precision	Recall	F-score	Support	Kappa
Base	0.023	0.062	0.033	64.062	-0.0023
SMOTE	**0.155**	**0.158**	**0.144**	**216.625**	**0.107**
Borderline-SMOTE	**0.148**	**0.190**	**0.144**	**190.188**	**0.159**
ROS	**0.271**	**0.345**	**0.287**	**216.625**	**0.305**
RUS	0.028	-0.109	0.038	4.812	0.004
SVC+W	**0.121**	0.068	0.069	64.062	0.003
SVC	0.035	0.066	0.040	64.062	0.016
DTC	0.049	0.047	0.048	64.062	-0.018
DTC+W	0.060	0.058	0.059	64.062	0.005
Logreg	0.125	0.069	0.064	64.062	-0.001
Logreg+W	0.062	0.052	0.035	64.062	-0.001
Ridge	0.041	0.062	0.046	64.062	0.001
Ridge+W	0.041	0.044	0.016	64.062	-0.007
MNB	0.061	0.058	0.047	64.062	-0.005
RFC	0.051	0.059	0.043	64.062	-0.008
RFC+W	0.068	0.063	0.045	64.062	0.005
XGBOOST	0.055	0.068	0.050	64.062	0.024
Bag	0.055	0.058	0.052	64.062	-0.008
Calibration SVC	0.035	0.063	0.028	64.062	0.004
Calibration RFC	0.033	0.063	0.026	64.062	0.003

Without additional regularisation, we got an average macro F1-score of 0.06 and an average macro recall of 0.08. Table IX gives the classification report for this training. With additional regularization, we got an average macro F1-score of 0.09 and an average macro recall of 0.09. Thus, we got slightly better results, but still, the results have much space for improvement. Table X gives the classification report for this training. Thus, additional regularization helps to improve the results of a classification in an imbalanced dataset.

Table IX LSTM only with dropout regularization

Class	Precision	Recall	F1-score	Support
INFP	0.00	0.00	0.00	20
INFJ	0.00	0.00	0.00	0
INTP	0.00	0.00	0.00	0
INTJ	0.00	0.00	0.00	0
ENTP	0.00	0.00	0.00	0
ENFP	0.80	0.27	0.40	60
ISTP	0.00	0.00	0.00	0
ISFP	0.00	0.00	0.00	0
ENTJ	0.00	0.00	0.00	0
ISTJ	0.00	0.00	0.00	0
ENFJ	0.21	0.95	0.34	20
ISFJ	0.00	0.00	0.00	0
ESTP	0.00	0.00	0.00	0
ESFP	0.40	0.10	0.16	20
ESFJ	0.00	0.00	0.00	0
ESTJ	0.00	0.00	0.00	20
accuracy				0.28
macro avg	0.09	0.08	0.06	140
weighted avg	0.43	0.26	0.24	140

In addition, we can conclude that pre-trained models and additional regularization can improve classification results on imbalanced datasets by comparing these results with base models.

Table X LSTM with additional L1/L2 regularization

Class	Precision	Recall	F1-score	Support
INFP	0.00	0.00	0.00	0
INFJ	0.00	0.00	0.00	0
INTP	0.00	0.00	0.00	0
INTJ	0.00	0.00	0.00	0
ENTP	0.00	0.00	0.00	0
ENFP	0.60	0.24	0.34	50
ISTP	0.00	0.00	0.00	25
ISFP	0.00	0.00	0.00	0
ENTJ	0.32	0.36	0.34	25
ISTJ	0.00	0.00	0.00	0
ENFJ	0.57	0.87	0.68	75
ISFJ	0.00	0.00	0.00	0
ESTP	0.00	0.00	0.00	0
ESFP	0.00	0.00	0.00	0
ESFJ	0.00	0.00	0.00	0
ESTJ	0.00	0.00	0.00	0
accuracy				0.25
macro avg	0.09	0.09	0.09	175
weighted avg	0.46	0.49	0.44	175

Regarding the semantical aspect of error, we evaluated semantic textual similarity among MBTI classes, and Figure 4 presents the results. Then, we computed the correlation between semantic similarity and Hamming distance between MBTI classes. The highest similarities have classes with the lowest Hamming distance between classes, as expected. As a result, we got a correlation coefficient of -0.516 and a small p-value that we can consider zero value. In addition, these results suggest that future research can take the direction of semantic similarity and Hamming distance between MBTI classes for further improvements of classification.

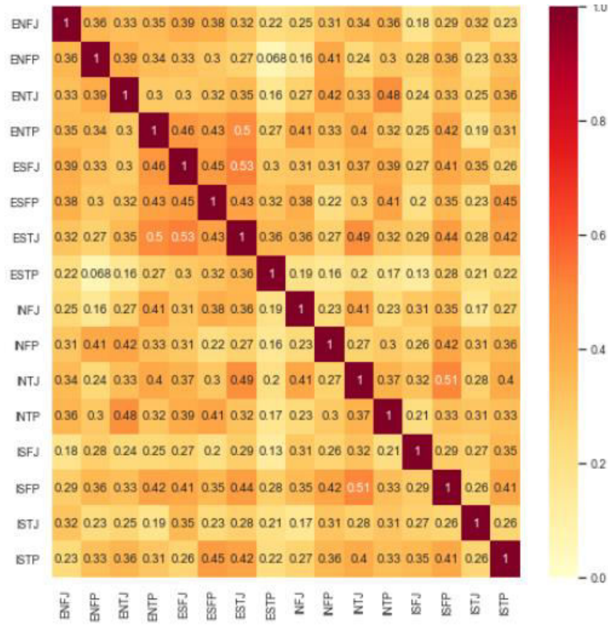

Figure 4. Semantic textual similarity for MBTI classes

V. CONCLUSION

Our research shows that in dealing with an imbalanced MBTI dataset as the NLP problem of multiclass classification, using pre-trained models, SMOTE, Borderline-SMOTE, and ROS approaches gave better results than the base model and other implemented methods in our research. However, among these, the ROS approach provided the best results.

These approaches significantly improve outcomes, especially in datasets with a high imbalance and many classes like our MBTI dataset. In addition, including semantic textual similarity can give a new perspective of relations between classes and help to improve MBTI classification.

For further researches, we plan to experiment with different architectures of neural networks and different ensemble methods. Also, we will try different approaches using semantic aspects of error.

VI. REFERENCES

[1] Y. LeCun, Y. Bengio and G. Hinton, "Deep Learning" in Nature, vol. 521., pp. 436-444, 2015.

[2] D. Liu, Y. Li and M. A. Thomas. "A Roadmap for Natural Language Processing Research in Information Systems." Hawaii International Conference on System Sciences. Honolulu, 2017.

[3] A. U. Rehman, A. K. Malik, B. Raza, and W. Ali, "A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis," in Multimedia Tools and Applications, 2019

[4] M. Ren, W. Zeng, B. Yang, R. Urtasun, "Learning to Reweight Examples for Robust Deep Learning," https://arxiv.org/pdf/1803.09050.pdf

[5] V. G. J. W. S. C. S. V. T. M. E. H. Steven Y. Feng, »A Survey of Data Augmentation Approaches for NLP,« ACL-IJCNLP 2021, Bangkok, 2021.

[6] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: "SMOTE: synthetic minority over-sampling technique," J. Artif. Intell. Res. 16, 321–357 (2002)

[7] K. Z. Jason Wei, »EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks,« Conference on Empirical Methods in Natural Language Processing, Hong Kong, 2019

[8] Saul Calderon-Ramirez, Shengxiang-Yang, Armaghan Moemeni, David Elizondo, Simon Colreavy-Donnelly. Correcting Data Imbalance for Semi-Supervised Covid-19 Detection Using X-ray Chest Images. Preprint: https://www.researchgate.net/publication/343759366_Correcting_Data_Imbalance_for_Semi-Supervised_Covid-19_Detection_Using_X-ray_Chest_Images, 2020.

[9] Veysel Kocaman, Ofer M. ShirThomas, BäckThomas Bäck. Improving Model Accuracy for Imbalanced Image Classification Tasks by Adding a Final Batch Normalization Layer: An Empirical Study. https://www.researchgate.net/publication/345788674_Improving_Model_Accuracy_for_Imbalanced_Image_Classification_Tasks_by_Adding_a_Final_Batch_Normalization_Layer_An_Empirical_Study: Preprint, 2020.

[10] Vipul Arora, Ming Sun, Chao Wang. "Deep Embeddings for Rate Audio Event Detection with Imbalanced Data." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019.

[11] C.G.Jung. Psychological Types (The Collected Works of C. G. Jung, Vol. 6). Princeton University Press, 1976.

[12] Isabel Briggs Myers, Peter B. Myers. Gifts Differing: Understanding Personality Type 2nd ed. CPP, 1995.

[13] T. M. &. B. Foundation, "The Myers & Briggs Foundation," 2021.[Online].Available: https://www.myersbriggs.org/my-mbti-personality-type/my-mbti-results/how-frequent-is-my-type.htm. [Accessed 08 07 2021].

[14] Mike Schuster, Kuldip K. Paliwal. "Bidirectional Recurrent Neural Networks." IEEE Transactions on Signal Processing 45.11 (November 1997): 2673-2681.

[15] Alex Gravesa, Ju¨rgen Schmidhuber. "Framewise phoneme classification with bidirectional LSTM and other neural network architectures." Neural Networks 18. Special Issue (2005): 602–610.

[16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting" in Journal of Machine Learning Research, vol. 15, pp. 1929-1958, 2014

[17] Diederik P. Kingma, Jimmy Lei Ba, "Adam: a Method for Stochastic Optimization," ICLR, 2015

[18] M. Frković, N. Čerkez, B. Vrdoljak and S. Skansi: "Evaluation of Structural Hyperparameters for Text Classification with LSTM Networks ", Mipro 2020

[19] Han, H., Wang, W.Y., Mao, B.H.: "Borderline–SMOTE: a new over-sampling method in imbalanced data sets learning. "In Proceedings of the 2005 International Conference on Intelligent Computing

[20] Justin M. Johnson, TaghiM. Khoshgoftaar: "The Effects of Data Sampling with Deep Learning and Highly Imbalanced Big Data, "Information Systems Frontiers, Springer 2020, https://doi.org/10.1007/s10796-020-10022-7