JUITA: Jurnal Informatika e-ISSN: 2579-8901; Vol. 12, No. 2, November 2024

Improving Sentiment Analysis Accuracy Using CRNN on Imbalanced Data: a Case Study of Indonesian National Football Coach

Slamet Riyadi[1*], Muhammad Dzaki Mubarok[2], Cahya Damarjati[3], Asnor Juraiza Ishak[4]

[1,2,3]*Information Technology Study Program, Faculty of Engineering, Universitas Muhammadiyah Yogyakarta, Indonesia*

[4]*Faculty of Engineering, Universiti Putra Malaysia, Malaysia*

*corr_author: riyadi@umy.ac.id

Abstract - Conducting sentiment research on the perception of the Indonesian people towards Shin Tae Yong's (STY) role as coach of the Indonesian National Football Team (PSSI) is crucial as it can assist PSSI in determining whether to extend STY's contract. Prior studies have demonstrated that Deep Learning achieves a high level of accuracy when applied to sentiment analysis in many domains. Nevertheless, no investigation has been conducted thus far utilizing deep learning techniques to examine emotion towards STY. This study employs modified Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Convolutional Recurrent Neural Networks (CRNN), and CRNN models with and without data oversampling. The research findings indicate that the CRNN model, when combined with data oversampling and a redesigned architecture, achieves the highest level of accuracy (1.00) and consistently performs well. This research provides significant contributions in three areas: firstly, it utilizes Deep Learning techniques for sentiment analysis on STY; secondly, it modifies the CRNN architecture; and thirdly, it applies data oversampling to address the issue of imbalanced data.

Keywords: sentiment analysis, Deep Learning, overfitting, oversampling, imbalanced data.

I. INTRODUCTION

Football is one of the most widely recognized sports worldwide. Around 200 million people play this sport, with 40 million of them being women [1]. Emerging in the 19th century, football has grown to become a sport played in almost every country, including Indonesia [2]. Football entered Indonesia in the early 20th century when the Dutch colonialists introduced it, and it has since become increasingly popular among the locals [3]. Football holds an extraordinary appeal, creating its own culture and intertwining it with various aspects such as fashion, politics, and more [4]. Moreover, football boasts passionate fan groups who deeply love the teams they support [5].

The advancement of technology in this era, especially with social media connecting people worldwide, has significantly impacted various aspects, including football [6], [7]. Social media serves as a platform for connecting football players and teams to their supporters. Typically, it provides a space for supporters to give suggestions or criticisms to players and teams through tweets or comments on players' personal accounts and official football team accounts. One example is the relationship between supporters and the Indonesian National Football Team. Since the appointment of Shin Tae-Yong, often referred to as STY, as the coach of the Indonesian National Team on December 28, 2019, many supporters have expressed their views on STY through social media. It is because the performance of the Indonesian National Team, whether good or bad, will impact the perception of supporters and the Indonesian community towards STY. Since STY previously led the South Korean National Team to victory against the German National Team in the 2018 FIFA World Cup, many Indonesian supporters have high hopes that STY will improve the Indonesian National Football Team.

The perspective of Indonesian supporters or society towards STY since his coaching era can be observed from tweets and comments on social media. Sentiment analysis can be used to determine the sentiment of tweets and comments regarding STY, where sentiment analysis is a Natural Language Processing (NLP) technique used to determine whether the emotions implied in the text are positive, negative, or neutral [8]. This is crucial, considering that supporters and society are vital for a football team, and if supporters do not like a coach or player, the football team may face consequences [9]. This could potentially happen to STY, regardless of his achievements in coaching other countries. That's why sentiment analysis regarding the views of Indonesian supporters and society towards STY is essential, so that the results can be considered by the Indonesian team in deciding whether to retain STY as a coach or not.

Sentiment analysis can be approached in various ways, one of which is using Deep Learning (DL) methods. DL is a sophisticated algorithm for understanding data. It mimics the human brain's way of processing data by examining it in multiple layers. Each layer of DL learns higher levels of data abstraction than the previous layer. It allows DL to understand complex non-linear data that traditional machine learning algorithms may struggle with [10].

In previous research using CNN for sentiment analysis of movie reviews and tweets, the highest accuracy obtained was 80.9% for movie review datasets and 81.1% for Twitter datasets. The advantage of this research is the improved accuracy of sentiment analysis, although it has not reached the accuracy achieved by other studies using more complex CNN architectures. It presents a gap in the research, where further development of more sophisticated CNN architectures is needed to handle larger datasets [11]. Another study conducted sentiment analysis on tweets and movie reviews using the RNN method, achieving an average accuracy of 80% in classifying positive and negative sentiments from the dataset. The advantage of this research is its ability to perform sentiment analysis on datasets using languages other than English. However, the study has not achieved higher accuracy, indicating the need for deeper datasets to improve accuracy [12].

Besides CNN and RNN, another method that can be used for sentiment analysis is the Convolutional Recurrent Neural Network (CRNN). In another study using CRNN for sentiment analysis, the highest accuracy obtained was 98%. This research used datasets in the Arabic language. The advantage of this study is that it achieves better results in sentiment classification tasks compared to other architectures. However, the limitation lies in the Arabic language dataset, which may have different structures than other languages, potentially affecting the CRNN model's performance when using datasets in different languages [13].

The current research will modify the CRNN architecture to achieve higher accuracy than previous studies. This research will use datasets from sentiment analysis of Indonesian society's views on STY, which have not been used in previous studies. This research will provide a comparison of CNN, RNN, CRNN, and modified CRNN for sentiment analysis of STY. This research is limited to using datasets in Indonesian. Testing with datasets in other languages has not been conducted and will be suggested for future research. This is important to test the generalization of the modified CRNN model across different languages and cultural contexts, making the results more applicable globally.

II. METHOD

A. Data Collection

This research utilizes data scraping techniques to gather comments or tweets directed towards STY. Data scraping is the process of manually or automatically extracting data from websites. This technique aims to efficiently collect data without having to inspect each one individually, which would be time-consuming and more complex [14].

Google allows developers to access the Application Programming Interface (API) using Google for Developers. This research utilizes Google for Developers to obtain the API from YouTube and uses the Google API Client, a library in Python, to scrape data from YouTube. We target comments on videos where STY is the speaker and which have been viewed at least 300,000 times.

Our dataset comes from comments on three YouTube videos titled "STY: Sacrifice For the Indonesian National Team, Set High Targets in the ASIAN CUP?", "PDP EPS 31 - WHAT'S UP IN 2024?? COACH SHIN AND BANG KIM KNOW THE ANSWER!!" and "EXCLUSIVE INTERVIEW STY: IF THE NATIONAL TEAM WANTS TO BE STRONG, IT NEEDS A STRONG LEAGUE!!". From these three videos, 5209 comments were collected, the details of which can be seen in Table I.

B. Data Preparation

Data preparation is a stage in which the dataset owned is prepared before moving to the next stage. The preprocessing stage is carried out to improve accuracy and eliminate noise and irrelevant data [15].

In this study, various text normalization and cleaning techniques were applied to ensure that the data used in sentiment analysis is clean, relevant, and consistent. First, converting text to lowercase was done to maintain text data consistency and reduce processing complexity. Emoji removal was then performed to eliminate symbols irrelevant to text-based analysis. Techniques for reducing letter repetition and removing word repetition were applied to ensure that the text does not have excessive characters or words that could disrupt the analysis. Additionally, removing extra whitespace was performed to keep the text neat and well-structured.

TABLE I
THE DATA OBTAINED

No.	Published	Comment	Like Count
1	2023-10-21T13:58:15Z	*Betul dulu menit 30 ama 75 udah ga ada tenaga apalagi kl pas kalah, malah ancurrrrr... jujur harus diakui di era STY stamina pemain timnas josss*	12
2	2023-10-21T15:15:34Z	*Harusnya di didik dari kecil lebih bagus, ngga harus pas masuk timnas baru di didik seperti sty jadi metalnya lebih bagus ngga sementaraðŸ™□ðŸ□» negara sepak bola yang maju dari kecil didikan sepakbolanya sangat kuat*	9
5207	2023-10-21T12:41:13Z	*melihat komposisi pemain timnas indonesia sangat menjajikan*	42

The technique of removing banned words was also applied to eliminate undesirable words, such as vulgarities or irrelevant abbreviations. Subsequently, a series of steps were taken to remove URLs, HTML tags, numbers, currency symbols, and punctuation marks to ensure the text is clean and free from non-text elements. Normalization using a slang dictionary was carried out to replace slang abbreviations with their more formal and standard equivalents, thereby enhancing the model's understanding of the text.

These techniques were chosen because each contributes to different aspects of text cleaning and normalization, from removing irrelevant elements to aligning text structure. The impact on model performance is significant; with cleaner and more consistent data, the model can learn from more meaningful text representations, reduce noise, and avoid biases that may arise from disorganized data. This ultimately improves the accuracy and generalization of the model for sentiment analysis tasks.

After the data preparation process, the dataset, which initially consisted of 5209 entries, was reduced to 5207 and will be transformed as shown in Table II.

C. Data Labelling

After data preparation, each piece of data will be labeled as either a positive, negative, or neutral sentiment. The process of labeling data is done using TextBlob, which is a Python library used for NLP. TextBlob supports multiple languages but can be challenging for some languages in the data labeling process [16]. Since the dataset is in Indonesian, before labeling, the prepared data is translated into English using a Python library called GoogleTrans, and then TextBlob will label each translated data as shown in Table III. After each piece of translated data is labeled, validation is carried out to determine whether the labeling results are correct or incorrect. From the validation results, there were some data points that were incorrectly labeled and corrected manually. After labeling the dataset, it will be divided into three parts: 80% for training data, 10% for testing data, and 10% for validation data [17].

D. Building The Model Architecture

In previous research [18], the CRNN architecture is shown in Fig. 1a. The input data will be processed using the CNN method and then processed using the RNN method. For clarity, the convolutional layers form a neural network that can minimize errors. The output is then sent to the dropout layer and combined to monitor the number of neurons in the previous layer. The max-pooling layer receives small blocks from the previous convolutional layer to achieve maximum output by reducing errors [19]. Furthermore, in the RNN method, the output from CNN will be sent and learned by LSTM. LSTM will remember each step of the previous actions and add its current state. Then, Softmax will receive the output value of the last model.

TABLE II
THE DATASET AFTER BEING PREPARED

No.	Comment	Clean_text	Normalization
1	*Betul dulu menit 30 ama 75 udah ga ada tenaga apalagi kl pas kalah, malah ancurrrrr...jujur harus diakui di era STY stamina pemain timnas josss*	*betul dulu menit ama udah ga ada tenaga apalagi kl pas kalah malah ancurr jujur harus diakui di era sty stamina pemain timnas joss*	*betul dahulu menit sama sudah tidak ada tenaga apalagi kalau pas kalah bahkan ancurr jujur harus diakui di era sty stamina pemain timnas joss*
2	*Harusnya di didik dari kecil lebih bagus, ngga harus pas masuk timnas baru di didik seperti sty jadi metalnya lebih bagus ngga sementaraðŸ™□ðŸ□» negara sepak bola yang maju dari kecil didikan sepakbolanya sangat kuat*	*harusnya di didik dari kecil lebih bagus ngga harus pas masuk timnas baru di didik seperti sty jadi metalnya lebih bagus ngga sementara negara sepak bola yang maju dari kecil didikan sepakbolanya sangat kuat*	*harusnya di didik dari kecil lebih bagus tidak harus pas masuk timnas baru di didik seperti sty jadi metalnya lebih bagus tidak sementara negara sepak bola yang maju dari kecil didikan sepakbolanya sangat kuat*
5205	*melihat komposisi pemain timnas indonesia sangat menjajikan*	*melihat komposisi pemain timnas indonesia sangat menjajikan*	*melihat komposisi pemain timnas indonesia sangat menjajikan*

TABLE III
THE LABELED DATA

No	translated_text	Subjectivity	Polarity	Sentiment
1	it's true that in the past, i didn t have any energy especially when i lost i was devastated honestly it has to be acknowledged that in this era the stamina of national team players joss	0.7	0.175	positive
2	should you be educated from a young age it s better don t you have to be educated when you enter the new national team like sty so the metal is better isn t it while football countries that have developed from childhood have a very strong football education	0.5179	0.3166	positive
5207	seeing the composition of the Indonesian national team players is very promising	0.65	0.26	positive

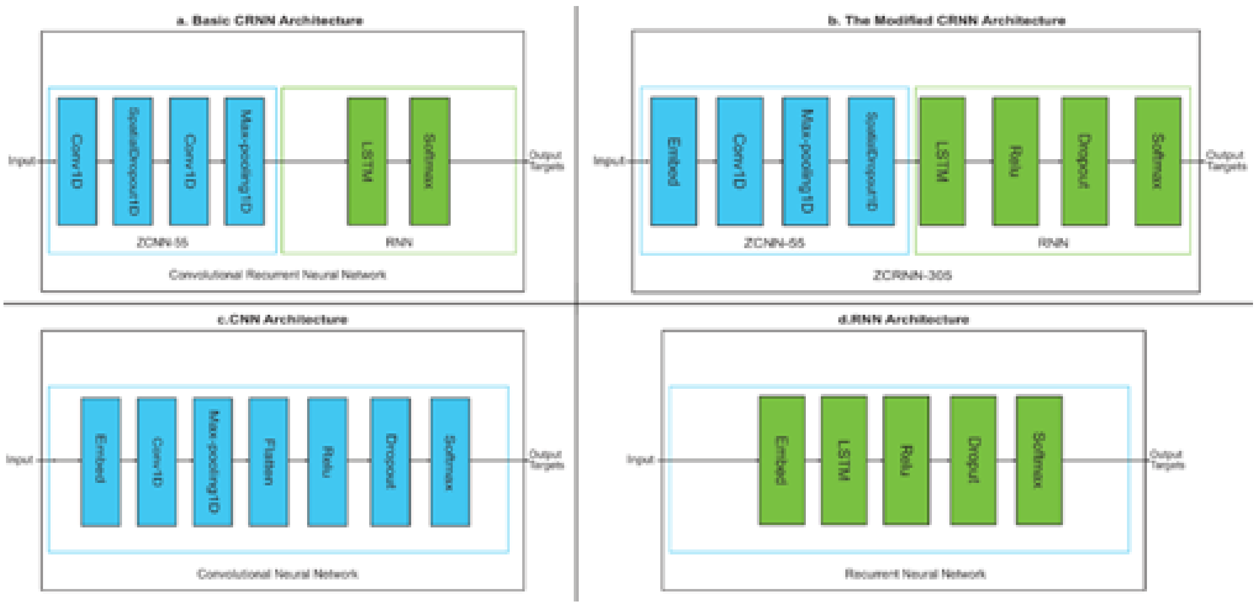

Fig. 1 (a) Basic CRNN architecture, (b) The modified CRNN architecture, (c) CNN architecture, (d) RNN architecture

This research will modify the CRNN architecture by adding or exchanging certain parts to improve the model's generalization ability, avoid overloading, enhance the model's performance for imbalanced classes, broaden the knowledge about generalization models, and reduce dilution and explosion problems. By allowing the model to learn data representations at different levels, model convergence is accelerated during training, enabling the model to learn new features from diverse data and enhancing gradient distribution, thus improving overall model performance [20]. The main modifications include the addition of an embedding layer, the repositioning of the spatial dropout layer, and the addition of a ReLU activation layer and an extra dropout layer. The architecture used in this research can be seen in Fig. 1b.

Sentiment analysis begins by converting text into numeric vector representations, and the embedding layer then modifies this representation into more meaningful vectors, capturing deep correlations between words in the text. After that, Conv1D and Max pooling1D are used to extract local and global patterns from the text, enhancing resilience to disturbances. To avoid overfitting, SpatialDropout1D is applied, while LSTM is used to understand long-term dependencies in the text. ReLU and Dropout are used to increase the complexity and generalization of the model. Ultimately, Softmax generates probabilities for each sentiment class, with the target output indicating the actual sentiment class.

In this study, two model architectures were also used to compare the performance of models in text sentiment analysis: RNN and CNN. The RNN model is designed to process sequences of words in text by leveraging its recurrent nature to capture complex temporal relationships. This is achieved using an Embedding layer as a representation of word vectors, followed by an LSTM layer for sequential processing of the sequences. The model then utilizes Dense layers for further feature learning and softmax output for multi-class sentiment classification. On the other hand, the CNN model employs Conv1D to extract local features from word embedding vectors, with MaxPooling1D to reduce data dimensions and flatten to reshape the output before entering Dense layers for classification. The architectures can be seen in Fig. 1c and Fig. 1d.

E. Training and Evaluating the Model

Before the model is trained, it will undergo a compilation process, which is a step that determines how the model will be optimized and evaluated. This process has parameters such as an optimizer, a loss function, and metrics. The optimizer helps find the optimal design by considering various criteria and prioritizing constraints, which are used to evaluate how closely the predicted results by the network match the predetermined truth labels [21]. Metrics are tools used to assess the performance of your model. The metric function and loss function are similar but not the same; the evaluation results of metrics are not used during model training. It is important to note that any loss function can be used as a metric [22].

After the model passes the compilation process, it enters the training process. This process takes the training data and uses parameters such as the number of epochs and batch size. Epoch is the number of iterations performed on the entire training data to train the model. This number should be optimal because it affects how well the model can understand patterns in the data and prevents overfitting [23]. Batch size is the number of training data samples used in each iteration of the DL training process, the size of which should be optimized to achieve the best results [24]. Besides epoch and batch size, there is also validation data in this process, which is used to evaluate the model's performance and can provide insight into how well the model processes data not used during the training process. It can help decide whether the model needs adjustments to improve its performance. In this study, several trials were conducted using the following parameters: 25 epochs with a batch size of 32, 50 epochs with a batch size of 64, and finally, the parameters were set to 100 epochs and a batch size of 64. Among these three parameter sets, the model performed optimally with 100 epochs and a batch size of 64, so this parameter set was used for model comparison.

Furthermore, the model will be evaluated using test data to test how well it can make predictions on new data. The output of this process is the evaluation result in the form of the accuracy value of the model. The evaluation process is also conducted using a Confusion Matrix, an evaluation tool to indicate the number of correct and incorrect predictions. We can obtain the accuracy, precision, recall, and f1-score values from the Confusion Matrix using the formulas seen in Fig. 2 [25], [26]. Accuracy measures the accuracy of correct predictions, precision measures the predictive value of a label, recall measures the algorithm's ability to detect true positive examples, and f1-score is the harmonic mean of precision and recall [27]. A comparison was made from the evaluation results of each model architecture tested, and the parameters used had to be the same to ensure a fair comparison.

III. RESULT AND DISCUSSION

A. Sentiment Analysis

In labeled datasets, certain words consistently appear in each comment. These words can be visualized using WordCloud, as shown in Fig. 3.

In the process of labeling each comment, there are three sentiments or classes: positive, negative, and neutral. The counts of these three classes will be calculated and compared with 5207 data points or comments. We can see the comparison of the three classes in a bar graph, as shown in Fig. 4.

Based on the graph shown in Fig. 4a, comments with a positive sentiment are higher than those with a count of 2284 comments. Neutral sentiment comments are slightly lower than positive ones, with a count of 2066 comments. Meanwhile, negative sentiment comments are at 736, which is significantly lower compared to the other two sentiments.

B. Application of Deep Learning Models

The training process is initiated after constructing the architecture of the DL models to be used. We will utilize the modified CRNN architecture, basic CRNN, basic CNN, and basic RNN models in this process. Observing the graph in Fig. 4a, the comparison between the count of negative sentiments and the other two is significantly disproportionate. It indicates an imbalanced data characteristic, which can lead to overfitting during the model training process [28]. To mitigate the risk of overfitting during training, we address the data imbalance by duplicating the number of negative comments and neutral comments to increase their count to match the number of positive sentiment comments, which is 2284 comments. This treatment is commonly referred to as data oversampling using the resample technique [29]. The resulting graph from the oversampled data can be seen in Fig. 4b.

With both imbalanced and balanced datasets, we trained the models using both data types. It was done to compare how data quality could affect the training process. The results of training each model on balanced and imbalanced data are displayed in the accuracy model graph, as shown in Fig. 5.

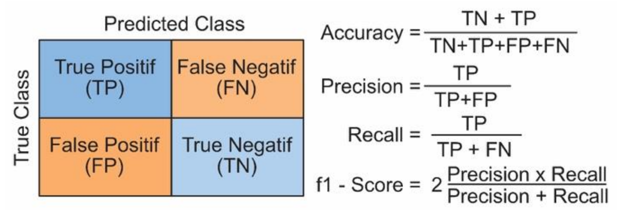

Fig. 2 Confusion Matrix and Its Formula

Fig. 3 Visualize frequently occurring words using a WordCloud

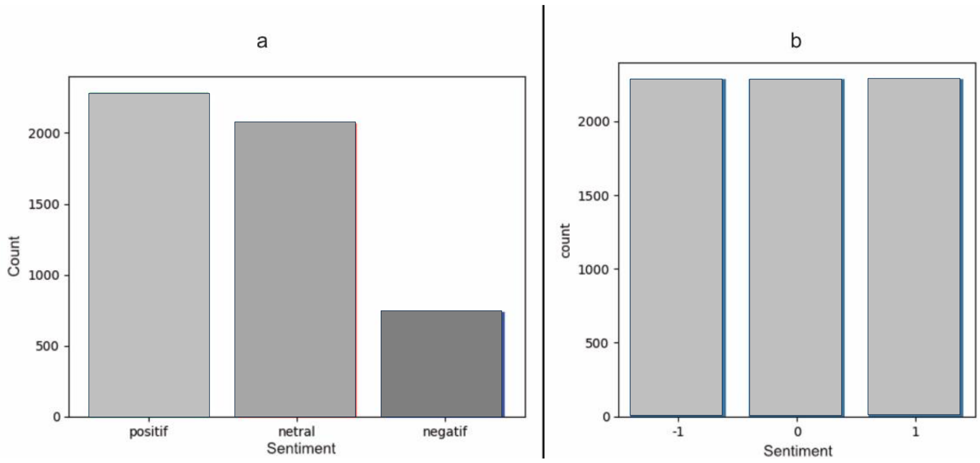

Fig. 4 (a) The sentiment distribution graph, (b) the graph of oversampled data results

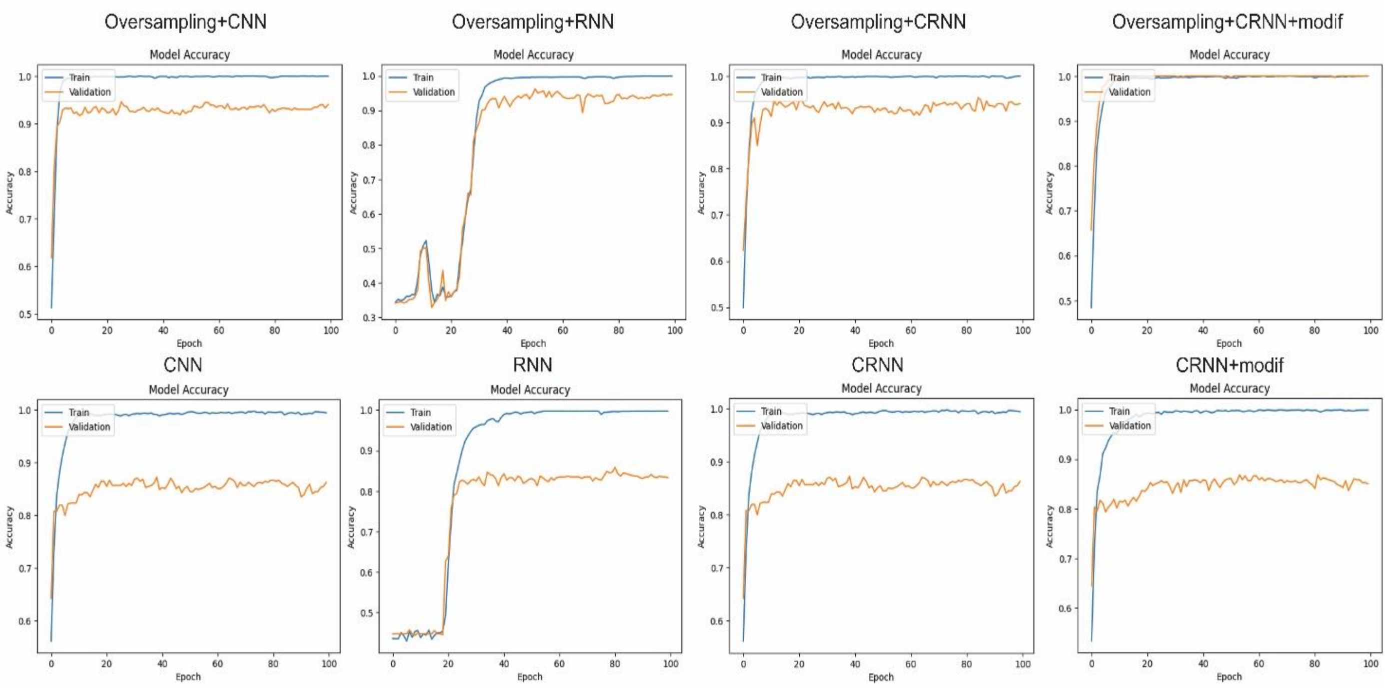

Fig. 5 The model accuracy graph for each model

In the eight graphs above, a significant difference is evident. The graphs of the models subjected to oversampling treatment show a smaller gap between the training and validation data lines than those without this treatment. The graphs without oversampling treatment exhibit signs of overfitting, with high accuracy on the training data but low accuracy on the validation data.

This indicates that the models are overly learning the training data patterns, leading to poor performance in new data [30]. The CRNN model, with modifications and oversampling, demonstrates a more stable performance than the others. It's also noticeable that the accuracy of the training and validation data of the CRNN model with modifications and oversampling is higher than in the other seven graphs.

After the training process, each model underwent evaluation using test data. The evaluation was conducted to determine how well the models could predict previously unseen data. This process calculates the model's accuracy and yields accuracy scores in percentage format. The accuracy of each of the eight models can be seen in Table IV.

Table IV shows that among the eight evaluated models, the model subjected to oversampling treatment achieved accuracy scores above 90%, even reaching 100% in the Oversampling+CRNN+Modification model. We can calculate accuracy, precision, recall, and F1-score from the Confusion Matrix, as seen in Table V.

TABLE IV
THE ACCURACY SCORES FOR EACH MODEL

Model	Accuracy	Model	Accuracy
CNN	86.25%	Oversampling +CNN	95.19%
RNN	83.89%	Oversampling +RNN	93.15%
CRNN	85.66%	Oversampling +CRNN	94.31%
CRNN+Modification	84.28%	Oversampling + CRNN+Modification	100%

TABLE V
COMPARISON OF CALCULATION RESULTS

Metode	Accuracy	Precision	Recall	F1-Score
Oversampling+CRNN+Modification	1.00	1.00	1.00	1.00
Oversampling+CRNN	0.94	0.93	0.99	0.96
Oversampling+RNN	0.93	0.95	0.97	0.96
Oversampling+CNN	0.95	0.95	0.98	0.96
CRNN+Modification	0.84	0.76	0.58	0.66
CRNN	0.86	0.79	0.63	0.70
RNN	0.84	0.66	0.54	0.59
CNN	0.86	0.74	0.59	0.66

Based on the comparison in Table V, the results of the Oversampling+CRNN+Modification model are superior to the others, with a value of 1.00 for accuracy, precision, recall, and F1-score. Therefore, the CRNN model with oversampled data and a modified architecture can achieve high accuracy.

IV. CONCLUSION

In this study, sentiment analysis was conducted on STY, the coach of the Indonesian national football team. Results revealed a prevalence of positive sentiments, indicating continued public trust in his leadership. DL techniques, including CNN, RNN, CRNN, and modified CRNN models, were employed for analysis due to their effectiveness in text classification. Oversampling was applied to prevent overfitting during model training and address the imbalanced dataset. Comparisons between models with and without oversampling showed clear benefits, with oversampled models consistently outperforming non-oversampled ones. The CRNN model with oversampled data and modified architecture demonstrated superior performance, achieving perfect accuracy, precision, recall, and F1 score. This underscores the importance of data treatment and model architecture modification in enhancing sentiment analysis accuracy for STY's comments.

ACKNOWLEDGEMENT

We would like to extend my sincere gratitude to Universitas Muhammadiyah Yogyakarta for their extraordinary financial support, which has played a significant role in the success of this research and publication. Without their assistance, this achievement would not have been possible.

REFERENCES

[1] J. Dvorak, A. Junge, T. Graf-Baumann, L. Peterson, and F-MARC FIFA Medical Assessment and Research Center, “Football is the most popular sport worldwide,” *Am. J. Sports Med.*, vol. 32, no. 1_suppl, pp. 3–4, Mar. 2004, doi: 10.1177/0363546503262283.

[2] B. Riffenburgh and B. Carroll, “THE BIRTH OF PRO FOOTBALL,” 1989.

[3] A. Fuller, “Approaching football in Indonesia,” *Soccer Soc.*, vol. 16, no. 1, pp. 140–148, Jan. 2015, doi: 10.1080/14660970.2014.954387.

[4] M. J. Power, P. Widdop, D. Parnell, J. Carr, and S. R. Millar, “Football and politics: the politics of football,” *Manag. Sport Leis.*, vol. 25, no. 1–2, pp. 1–5, Mar. 2020, doi: 10.1080/23750472.2020.1723437.

[5] I. Syahputra, “Terbentuknya Identitas Fans Sepak Bola sebagai Budaya Massa dalam Industri Media,” *INFORMASI*, vol. 46, no. 2, p. 205, Dec. 2016, doi: 10.21831/informasi.v46i2.11377.

[6] T. Aichner, “Football clubs’ social media use and user engagement,” *Mark. Intell. Plan.*, vol. 37, no. 3, pp. 242–257, May 2019, doi: 10.1108/MIP-05-2018-0155.

[7] W. Akram, Department of Computer Applications, GDC Mendhar, Poonch, India, R. Kumar, and Department of Computer Applications, GDC Mendhar, Poonch, India, “A Study on Positive and Negative Effects of Social Media on Society,” *Int. J. Comput. Sci. Eng.*, vol. 5, no. 10, pp. 351–354, Oct. 2017, doi: 10.26438/ijcse/v5i10.351354.

[8] F. Aftab, S. U. Bazai, S. Marjan, L. Baloch, S. Aslam, A. Amphawan, and T. K. Neo, “A Comprehensive Survey on Sentiment Analysis Techniques,” *Int. J. Technol.*, vol. 14, no. 6, p. 1288, Oct. 2023, doi: 10.14716/ijtech.v14i6.6632.

[9] B. Constandt, M. M. Parent, and A. Willem, “Does it really matter? A study on soccer fans’ perceptions of ethical leadership and their role as ‘stakeowners,’” *Sport Manag. Rev.*, vol. 23, no. 3, pp. 374–386, Jul. 2020, doi: 10.1016/j.smr.2019.04.003.

[10] Z. Hao, “Deep learning review and discussion of its future development,” *MATEC Web Conf.*, vol. 277, p. 02035, 2019, doi: 10.1051/matecconf/201927702035.

[11] Moch. A. Nasichuddin, T. B. Adji, and W. Widyawan, “Performance Improvement Using CNN for Sentiment Analysis,” *IJITEE Int. J. Inf. Technol. Electr. Eng.*, vol. 2, no. 1, Jul. 2018, doi: 10.22146/ijitee.36642.

[12] M. Thomas and L. C.A, “Sentimental analysis using recurrent neural network,” *Int. J. Eng. Technol.*, vol. 7, no. 2.27, p. 88, Aug. 2018, doi: 10.14419/ijet.v7i2.27.12635.

[13] A. Onan, “Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 5, pp. 2098–2117, May 2022, doi: 10.1016/j.jksuci.2022.02.025.

[14] I. S. H. Almaqbali, “Web Scrapping: Data Extraction from Websites,” Jul. 2020.

[15] E. Haddi, X. Liu, and Y. Shi, “The Role of Text Pre-processing in Sentiment Analysis,” *Procedia Comput. Sci.*, vol. 17, pp. 26–32, 2013, doi: 10.1016/j.procs.2013.05.005.

[16] S. Dewi and D. B. Arianto, “TWITTER SENTIMENT ANALYSIS TOWARDS QATAR AS HOST OF THE 2022 WORLD CUP USING TEXTBLOB,” 2022.

[17] V. R. Joseph, “Optimal Ratio for Data Splitting,” *Stat. Anal. Data Min. ASA Data Sci. J.*, vol. 15, no. 4, pp. 531–538, Aug. 2022, doi: 10.1002/sam.11583.

[18] A. S. U., F. R. P. P., A. Abraham, and D. Stephen, “Deep Learning-Based BoVW–CRNN Model for Lung Tumor Detection in Nano-Segmented CT Images,” *Electronics*, vol. 12, no. 1, p. 14, Dec. 2022, doi: 10.3390/electronics12010014.

[19] K. Suzuki, S. G. Armato, F. Li, S. Sone, and K. Doi, “Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography,” *Med. Phys.*, vol. 30, no. 7, pp. 1602–1617, Jul. 2003, doi: 10.1118/1.1580485.

[20] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J. Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[21] B. Soujanya and T. Sitamahalakshmi, “Optimization with ADAM and RMSprop in Convolution neural Network (CNN): A Case study for Telugu Handwritten Characters,” *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 9, pp. 5116–5121, Sep. 2020, doi: 10.30534/ijeter/2020/38892020.

[22] K. Team, “Keras documentation: Metrics.” Accessed: Mar. 28, 2024. [Online]. Available: https://keras.io/api/metrics/

[23] X. Wu and J. Liu, “A New Early Stopping Algorithm for Improving Neural Network Generalization,” in *2009 Second International Conference on Intelligent Computation Technology and Automation*, Changsha, Hunan, China: IEEE, 2009, pp. 15–18. doi: 10.1109/ICICTA.2009.11.

[24] N. Rochmawati, H. B. Hidayati, Y. Yamasari, H. P. A. Tjahyaningtijas, W. Yustanti, and A. Prihanto, “Analisa Learning Rate dan Batch Size pada Klasifikasi Covid Menggunakan Deep Learning dengan Optimizer Adam,” *J. Inf. Eng. Educ. Technol.*, vol. 5, no. 2, pp. 44–48, Dec. 2021, doi: 10.26740/jieet.v5n2.p44-48.

[25] M. Artur, “Review the performance of the Bernoulli Naïve Bayes Classifier in Intrusion Detection Systems using Recursive Feature Elimination with Cross-validated selection of the best number of features,” *Procedia Comput. Sci.*, vol. 190, pp. 564–570, 2021, doi: 10.1016/j.procs.2021.06.066.

[26] A. H. Wicaksono, A. A. Supianto, S. H. Wijoyo, D. Krisnandi, and A. Heryana, “Klasifikasi Siswa Slow Learner untuk Mendukung Sekolah dalam Meningkatkan Pemahaman Siswa Menggunakan Algoritma Naïve Bayes,” *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 9, no. 3, pp. 589–596, Jun. 2022, doi: 10.25126/jtiik.2022935609.

[27] M. Sokolova, N. Japkowicz, and S. Szpakowicz, “Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation,” in *AI 2006: Advances in Artificial Intelligence*, vol. 4304, A. Sattar and B. Kang, Eds., in Lecture Notes in Computer Science, vol. 4304. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1015–1021. doi: 10.1007/11941439_114.

[28] Y. Sun, A. K. C. Wong, and M. S. Kamel, “CLASSIFICATION OF IMBALANCED DATA: A REVIEW,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 04, pp. 687–719, Jun. 2009, doi: 10.1142/S0218001409007326.

[29] R. Mohammed, J. Rawashdeh, and M. Abdullah, “Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results,” in *2020 11th International Conference on Information and Communication Systems (ICICS)*, Irbid, Jordan: IEEE, Apr. 2020, pp. 243–248. doi: 10.1109/ICICS49469.2020.239556.

[30] S. Pavlitskaya, J. Oswald, and J. M. Zöllner, “Measuring Overfitting in Convolutional Neural Networks using Adversarial Perturbations and Label Noise.” arXiv, Sep. 27, 2022. Accessed: Mar. 31, 2024. [Online]. Available: http://arxiv.org/abs/2209.13382