Emotion Intensity Value Prediction with Machine Learning Approach on Twitter

Rindy Claudia Setiawan[1*] and Andry Chowanda[2]
[1]Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University
Jakarta 11480, Indonesia
[2]Computer Science Department, School of Computer Science, Bina Nusantara University
Jakarta 11480, Indonesia
Email: [1]Rindy.claudia@binus.ac.id, [2]achowanda@binus.edu

***Abstract*—Recognizing the intensity of the emotions is a paramount task for an affective system. By recognizing the intensity of the emotions, the system can have better human-computer interaction. The research explores several machine learning approaches with several different feature extraction method combinations to solve the emotion intensity prediction task while also analyzing and comparing it with several previous related papers. The research uses the dataset provided through the WASSA 2017 and SemEval 2018 competition. The dataset utilizes four of the eight basic emotions that Plutchik defines (anger, fear, joy, and sadness). The total data result in 19,736 rows of entry, with a total of 10,715 (54.3%) for training, 1,811 (9.17%) for validation, and 7,210 (36.53%) for testing. Three feature extraction methods are used and compared: N-gram, TF-IDF, and Bag-of-Words. Meanwhile, machine learning algorithms are Linear Regression, Ridge Regression, K-Nearest Neighbor for Regression, Regression Tree, and Support Vector Regression (SVR). The results show that SVR with TF-IDF features has the best result of all attempted experiments, with a Pearson correlation score of 0.755 for all data and 0.647 for gold labels data. The final model also accepts newly seen data and displays the corresponding emotion label and intensity.**

***Index Terms*—Emotion Intensity, Value Prediction, Machine Learning Approach, Twitter**

I. INTRODUCTION

WORDS carry emotions within them. Generally, almost every sentence, especially in a first-person narrative, contains at least one emotion-related word to convey the speaker's intention and meaning precisely. Sometimes, these words may be obvious, such as unhappy, elated, or horrifying. However, it is not rare to find some examples where the sentence only insinuates the actual feeling through irony or sarcasm. It is why affective computing remains a very interesting area to research. Affective computing is a research field that combines various research domains, such as computer science, psychology, and cognitive science [1]. The purpose and target of this area of research are to enable emotionally challenged machines to react appropriately upon encountering human interaction or, more specifically, human affective states (e.g., moods and emotions). In this rising era of artificial intelligence systems, there have been numerous implementations in different subjects, such as education [2], healthcare [3], and business [4].

The rise of social media usage also benefits from the research. In the case of microblog applications like Twitter, sentiment analysis and emotion classification can also distill how the public reacts [5] to specific topics through hashtags, mentions, and topic-related words. Previous research presents the concept of emotion intensity [6]. Emotion intensity is an excellent complement to standard sentiment analysis systems. This additional information functions nicely as a complementary feature. In real-life uses, it helps in scaling or ranking the input. For example, when promoting a product, the reviews that exhibit joy more intensely can be highlighted more and reversely.

In the case of a customer service system, a priority scale based on the customers' degree of anger can also be generated. This practice may also benefit from other real-world uses, such as text-to-speech tasks and highlighting specific emotion tasks when distilling public response. This new concept touches upon a refreshingly new task in text analysis, which is regression. Previously the trends in text analysis mainly dealt with classification, which primarily employs artificial intelligence to classify emotions into a defined or discrete set of emotion labels.

However, this task is a regression task that differs

Received: May 10, 2022; received in revised form: July 22, 2022; accepted: July 22, 2022; available online: Sept. 18, 2023.
*Corresponding Author

from its predecessor. Regression basically predicts a continuously variable output. The only discrete value is the upper and lower bound, commonly set as between 0 and 1, although it may differ case by case.

Previous research that employs a regression model with a text-based dataset is very scarce and uncommon in nature. Thus, this consideration also adds a novelty value to this task. Despite that, up until now, there has not been any research nor model that can be mentioned as a benchmark for emotion-intensity topics that has received proper professional acclaim, except for the basic system created as a baseline for the competition [7] that also issues an emotion intensity dataset for analysis. The research on emotion intensity is still in its developing stages with limited resources available. The competition entry models are unpublished, and the original authors' benchmark models are relatively simple. Therefore, it leaves much room for personal modifications. Furthermore, publicly available datasets for emotion-intensity tasks are scarce, which also poses another problem, which is the lack of data compared to other Natural Language Processing (NLP) tasks.

The inspiration for the research comes from two competitions, WASSA 2017 [7] and SemEval 2018 [8]. The task defined in both competitions is to predict the intensity level of the emotion given the data of Twitter text and the corresponding emotion label. Since both competitions have ended, all training, development, and testing data have been provided. As stated by their developers, the first-ranking entry in the WASSA competition [9] may benefit from further exploring various feature extraction methods. The system consists of an ensemble of three broad sets of approaches combined using a weighted average of the separate predictions [9]. The approaches rely on using the word2vec approach as the feature extraction method and use Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) models to predict the intensity. However, they have to train different models for different emotion categories, and the basic models they used treated all words equally in the modeling of sentences. The second placing entry, SeerNet [10], implements an ensemble model that averages the top two models of the various models they attempt. They combine several lexicon models, Emoji embeddings, Glove embeddings, and Edinburgh embeddings, and mix-match them to obtain several combinations for the regressor. They also use several models: Support Vector Regression (SVR), AdaBoost, Random Forest, and Bagging Regressor.

Another approach is proposed for this task using machine and deep learning [11]. It averages both SVR and Bi-directional LSTM (BiLSTM) approach that utilizes N-gram and Tweet word embeddings for the features, respectively. Unlike more traditional approaches that use unidirectional language models to learn general language representations, Bidirectional Encoder Representations from Transformers (BERT) take a new approach which is the bidirectional model [12]. It results in a more robust token incorporating context from both directions. Due to this nature, BERT has recently become a state-of-the-art algorithm that can solve a wide array of NLP tasks.

Based on this analysis, methods with the Deep Learning (DL) approach tend to be higher than machine learning approaches. However, the machine learning implementation is not thoroughly explored and leaves much to be improved. Thus, the research thoroughly explore the machine learning implementation with further fine-tuning and parameter experimentation. A direct implementation of the pre-trained BERT model is also attempted for further analysis. Finally, the research proposes a system that can detect and rate the intensity and emotion of the input based on the extracted features. It aims to provide a framework that can efficiently do tasks with decent accuracy.

II. Research Method

The model proposed in the research consists of a regressor to predict the intensity of the underlying emotion based on the given text (i.e., tweets). The researchers try several combinations of well-known machine learning methods as the regressor base to find the best combination. The input consists of only text data, which are preprocessed into tokens and converted into features. The model is trained individually based on each emotion label. The overview flow of the experiment is visualized in Fig. 1.

A. Dataset Preprocessing

The researchers clean the unnecessary part of the texts before moving on to feature extraction to optimize the dataset for training. There are two different approaches to the preprocessing method:

1) Attempted method 1: remove stop words, convert them to lowercase, remove URLs, and remove the usernames mentioned.
2) Attempted method 2: remove stop words, convert them to lowercase, remove URLs, remove @ from the username mentioned, and remove # from hashtags.

These preprocessing methods are applied to training and validation datasets during the preliminary experiment. Based on several attempts at training data, method 1 works relatively better compared to method 2. Thus, further development uses method 1.

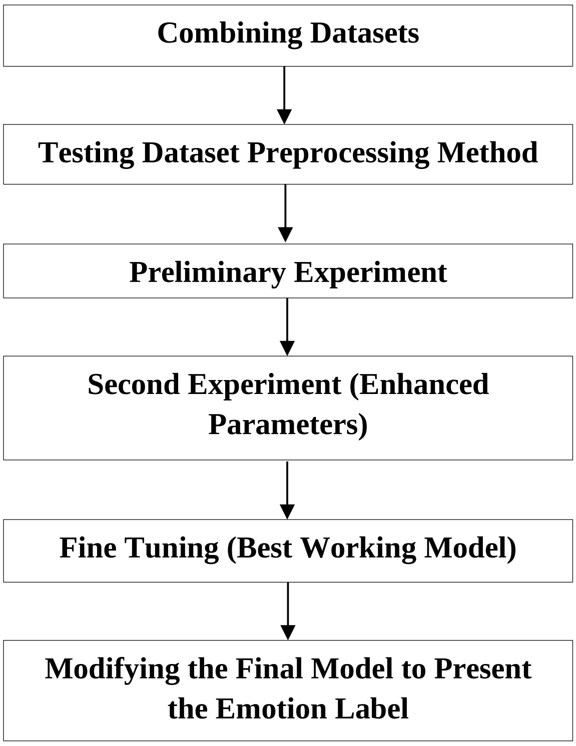

Fig. 1. The overview of the experiment working flow.

B. Preliminary Experiment

For feature extraction, the researchers try out three different methods: N-gram [13], TF-IDF [14], and Bag-of-Words [15]. Those methods are popular feature extraction for machine learning. The extracted vectors are forwarded to the classifier. The preliminary experiment uses a max feature of 1,000 with an N-gram range from 1 to 3. The preliminary setup for the regressor consists of no specific parameters for all algorithms: Linear Regression, Ridge Regression, Knn Regression (number of neighbors = 5), Decision Tree Regressor, and SVR.

The models are created using the sklearn [16] library in Python. The model is trained on the training dataset and tested on the validation dataset. While it is expected that the result may not be optimal, this phase is only conducted to get the first result data before the improvement phase. Based on this, the researchers make the first preliminary result on testing the validation dataset.

For the whole phase, the regression model consists of four parts. It is for each emotion: joy, fear, sadness, and anger. Thus, each part is retrained individually with the same method using training and validation datasets and used for testing on the test dataset.

A similar flow of work is implemented for BERT development: preprocessing, training, and testing. However, for the dataset, after preprocessing, the researchers convert it into Dataset class and DataLoader according to a common Pytorch setup. Then, for the main BERT model, the transformers library pre-trained ’bert-base-uncased’ variant is used with AdamW optimizer, 1e-5 learning rate, and 1e-8 epsilon value. The initial experiment used three epochs as recommended in the BERT paper, but due to lacking results, the researchers raise the epoch to 20 to see whether the result improves.

TABLE I
GRID SEARCH PARAMETER.

C	0.01; 0.1; 1; 10; 100; 1,000
Gamma	1; 0.1; 0.01; 0.001; 0.0001; “scale”
Tol	0.001; 0.0001; 0.00001

C. Second Experiment

Next, the experiment consists of enhancing and fine-tuning the best working combination based on the first result. In this step, the researchers modify the feature extractors’ parameters to improve the result. The max features for the feature extraction method are raised to 6,000 with an N-gram range from 1 to 4. The selection of the value 6,000 is because the total extractable feature from the dataset is 12,159. Therefore, the researchers select half and round it to the nearest thousand. This consideration is chosen due to the huge amount of conjunctions and interjections used in several entries, which may not be relevant to or reflect the entirety of the dataset.

D. Fine-Tuning

After obtaining the best working model from both the preliminary and new experiments, the next step involves fine-tuning the hyperparameter for the best working model. Grid Search [17] is applied for this step to optimize the results. The attempted parameter range is shown in Table I. The initial parameter for SVR based on the sklearn default setting is C: 1, gamma: ”scale”, and tol: 0.001. C refers to the regularization parameter, gamma is the kernel coefficient, and tol is short for tolerance for stopping criteria.

E. Datasets

The researchers utilize the dataset that is also used in WASSA 2017 [7] and SemEval 2018 [8]. The dataset is taken from the publicly available website (http://saifmohammad.com/). It is the website of the competition organizer and the author’s papers. The

Cite this article as: R. C. Setiawan and A. Chowanda, “Emotion Intensity Value Prediction with Machine Learning Approach on Twitter”, CommIT Journal 17(2), 235–243, 2023.

TABLE II
DATASET STRUCTURE EXAMPLE.

ID	Text	Label	Value
31064	I’m so excited to see Nat tonight 😍😍.. And how happy and cheery she is! & then I’m even more excited for her to get on social media 😍 #BB18”	Joy	0.979

TABLE III
SHR FOR THE EMOTIONS IN WASSA 2017 DATASET.

Emotion	Spearman	Pearson
Anger	0.779	0.797
Fear	0.845	0.850
Joy	0.881	0.882
Sadness	0.847	0.847

research combines both dataset as a singular dataset and retained the original process and emotion labels. This dataset utilizes four of the eight basic emotions defined by Plutchik [18]. It consists of anger, fear, joy, and sadness. An example of the data is shown in Table II.

The reliability of the final data is assessed through average Split-Half Reliability (SHR). It splits the item into two halves and calculates the correlation between them. A higher correlation denotes the better quality of the dataset. This process is repeated 100 times, and the correlation across the two sets of rankings and intensity scores is averaged. The SHR data are shown in Table III.

One of the specific subtasks stated in the dataset is the importance of gold-labeled data. It is the instances with original intensity scores ≥ 0.5. It is because, in some applications, moderate or strong emotion is more relevant than slightly felt emotion. Thus, correctly predicting a high-scale intensity provides more meaning compared to correctly determining emotion intensity in the lower range of the scale. The researchers mainly use the Pearson correlation coefficient in evaluating the model with two different comparisons: all data and gold label (data with original intensity label $\geq$ 0.5), following the original metric from the issuer of the dataset.

III. RESULTS AND DISCUSSION

The preliminary experiments implement all features with the parameter of the max feature of 1,000. Table IV shows the result of the tests. The best result is obtained from SVR using the TF-IDF method with a 0.736 Pearson correlation value, 0.612 Pearson correlation value on gold label data, and 0.52 R^2 value. SVR method with Bag-of-Words (BoW) and N-gram comes in the second and third, respectively. This trend shows that SVR performs better compared to other regression algorithms on every single evaluation metric used. Moreover, the Decision Tree and Ridge methods also show comparable performance right below the SVR with a Pearson value of around 0.6. The BERT method, known as the state-of-art, unexpectedly performs way worse than the other methods. It can be due to the lack of fine-tuning on the model and the unsuitableness of using the raw model without any modifications directly on the data. Further exploration has been done on the BERT model, such as increasing the number of epochs, manipulating the learning rate and epsilon, freezing all except the last layer, and more. However, none gives significant improvement, and the result is still lower than the machine learning models.

TABLE IV
PRELIMINARY RESULTS.

Framework	Pearson	Gold Pearson	R^2
N-gram LinReg	0.5800	0.433	0.187
N-gram Ridge	0.6360	0.482	0.372
N-gram Knn	0.4970	0.363	0.200
N-gram Tree	0.6550	0.474	0.326
N-gram SVR	0.7160	0.576	0.504
TF-IDF LinReg	0.5890	0.432	0.174
TF-IDF Ridge	0.6540	0.495	0.425
TF-IDF Knn	0.4030	0.271	0.031
TF-IDF Tree	0.6500	0.482	0.309
TF-IDF SVR	0.7360	0.612	0.520
BoW LinReg	0.5820	0.432	0.160
BoW Ridge	0.6370	0.484	0.370
BoW Knn	0.4890	0.362	0.182
BoW Tree	0.6540	0.477	0.321
BoW SVR	0.7170	0.581	0.504
BERT	0.1290	0.014	-0.325

TABLE V
DATA DISTRIBUTION.

Dataset	Process	Anger	Fear	Joy	Sadness	Total
WASSA -2017	Train	857	1,147	823	786	19,736
	Validation	84	110	79	74	
	Test	760	995	714	673	
SemEval -2018	Train	1,701	2,252	1,616	1,533	
	Validation	388	389	290	397	
	Test	1,002	986	1,105	975	
Total		4,792	5,879	4,627	4,438	

The annotation method used is Best-Worst Scaling (BWS). The total data distribution for the datasets is visualized in Table V. Gold labels refer to data entries with an intensity value greater than or equal to 0.5. The total data resulted in 19,736 rows of entries, with a total of 10,715 rows (54,3%) for training, 1,811 rows (9,17%) for validation, and 7,210 rows (36,53%) for testing.

The following research step is fine-tuning the model using Grid Search on the best working model, TF-IDF

TABLE VI
OPTIMAL PARAMETER USING GRID SEARCH RESULTS.

	C	Gamma	Tol	Kernel
Anger	100	0.1	1 E-05	RBF
Fear	1000	0.1	1 E-05	RBF
Joy	100	0.1	1 E-05	RBF
Sadness	1,000	0.1	1 E-05	RBF

SVR, to find the best parameters for each emotion. The researchers pick several parameters to test: C: [0.01, 0.1, 1, 10, 100, 1000], gamma: [1, 0.1, 0.01, 0.001, 0.0001,"scale"], and tol: [0.001, 1e-04, 1e-05]. The result of the search is shown in Table VI. These parameters are used for another comparison of the fine-tuned model and the original model.

The next step involves modifying the max features parameter. During this phase, the researchers raise it to 6,000, almost half of the total parameters of the data, which are 12,159. The numbers of N-gram limit are also raised from (1, 3) into (1, 4) based on the considerations on the configuration of previous related works with similar methods. The result of this experiment is in Table VII. Generally, all method outcomes are slightly improved through this change with the tradeoff of the exponential rise of training time. However, the reverse happens for some algorithms, such as N-gram Linear Regression. The Pearson correlation value drops rather significantly from 0.58 to 0.309. It shows that raising the number of features does not necessarily improve the result for all methods. It is also interesting to note that the best working model, TF-IDF SVR, does not benefit much from this change. The 0.019 raise comes with the price of a 500% increase in training time. Thus, it can be concluded that a further increase in several parameters slightly improves the result by a minuscule value – not overly significant to the result. However, the increase in the training time may be overly time-consuming.

Then, the result of the applied optimal parameter is shown in Table VIII. The result decreases a little bit, from 0.755 to 0.725. The initial SVR uses a uniformed configuration for the hyperparameters, C: 1, Gamma: 'scale', Tol: 1e-3, and RBF kernel. Changing the hyperparameter causes overfitting on the dataset, which affects this drop. Thus, for the final framework of the model, the researchers keep the original SVR, which obtains optimal results for most emotions.

As seen from Table IX, the proposed method performs better compared to the result of some previous works. For this case, the researchers conduct some comparison with the first ranked entry in WASSA 2017, Prayas [9], that utilizes an ensemble of Feed-Forward Neural Networks, Multi-Task Deep Learning and Sequence Modeling using CNNs and LSTMs. While the Pearson correlation value only improves slightly from Prayas, the Pearson gold label improves quite a good bit by 0.076 points. In a real-life implementation, the gold label is arguably what matters the most when using a sentiment or emotion analysis model. The researchers need to analyze the data related to the emotion label, or in emotion intensity case, a strongly felt emotion represented with an intensity value of more than 0.5.

It has a decent result comparable with previous research by performing better in fear, sadness, joy, and all the gold label value. However, the result shows the worst result in predicting anger out of all the four emotion labels. It is also the same case as the SHR result from the dataset, which calculates the manual annotation reliability. It shows anger with the lowest score, while fear, joy, and sadness have relatively similar values. There are also unexpected findings that show most machine learning-based methods (Baseline, NUIG, and the research) show a lower correlation value in anger. In contrast, the deep learning-based method (SeerNet and Prayas) shows a high correlation value in anger.

Figure A1 in Appendix plots the actual intensity with the predicted intensity. Most points have aligned with the correct position, same actual intensity (x-value) and predicted intensity (y-value). In lower intensity and extremely high intensity, there is rarely an outlier that is falsely predicted. However, the model is still lacking in dealing with medium-intensity emotions, especially in the range of 0.3–0.7.

It is also stated earlier that the final system accepts newly seen data and can predict the correlating emotion label and intensity. A sample of this function is shown in Fig. 2. The system is modified to calculate the probability of every single emotion and show only the most significant emotion label alongside the predicted intensity. Based on several casual attempts made, the system manages to classify the predicted emotion label correctly. Even when the sentence is seemingly quite neutral (not showing any significant emotion), the system can assign low-intensity levels that are rather close in number for all emotion labels. On the other hand, the middle-intensity part (around 0.4–0.8) is hard to measure precisely without further proper annotation methods. However, the predicted intensity is more or less quite acceptable through approximation.

IV. CONCLUSION

The research explores several machine learning models to solve the emotion intensity prediction task. In particular, the research tries several feature extraction methods combinations and compares and analyzes

TABLE VII
RAISED PARAMETER RESULTS.

Framework	Pearson					Gold Pearson				
	Average	Anger	Fear	Joy	Sadness	Average	Anger	Fear	Joy	Sadness
N-gram LinReg	0.309	0.354	0.168	0.361	0.353	0.186	0.209	0.155	0.204	0.176
N-gram Ridge	0.695	0.682	0.701	0.698	0.700	0.525	0.531	0.496	0.513	0.559
N-gram Knn	0.499	0.478	0.539	0.447	0.534	0.361	0.385	0.406	0.307	0.347
N-gram Tree	0.655	0.617	0.746	0.622	0.635	0.458	0.418	0.539	0.406	0.467
N-gram SVR	0.730	0.696	0.764	0.737	0.722	0.591	0.572	0.608	0.585	0.600
TF-IDF LinReg	0.601	0.574	0.634	0.596	0.599	0.447	0.429	0.461	0.417	0.481
TF-IDF Ridge	0.731	0.715	0.757	0.727	0.723	0.577	0.588	0.588	0.532	0.598
TF-IDF Knn	0.511	0.479	0.542	0.493	0.532	0.360	0.338	0.393	0.337	0.373
TF-IDF Tree	0.662	0.598	0.742	0.611	0.699	0.477	0.374	0.568	0.426	0.540
TF-IDF SVR	0.755	0.710	0.801	0.758	0.753	0.647	0.595	0.694	0.625	0.673
BoW LinReg	0.543	0.560	0.555	0.508	0.551	0.402	0.416	0.413	0.341	0.437
BoW Ridge	0.734	0.704	0.775	0.738	0.720	0.576	0.555	0.599	0.562	0.590
BoW Knn	0.479	0.440	0.495	0.469	0.514	0.355	0.320	0.394	0.341	0.367
BoW Tree	0.673	0.649	0.739	0.651	0.655	0.488	0.442	0.566	0.457	0.489
BoW SVR	0.741	0.696	0.787	0.745	0.737	0.623	0.578	0.664	0.605	0.642

TABLE VIII
RESULT COMPARISON USING GRID SEARCH.

Framework	Pearson					Gold Pearson				
	Average	Anger	Fear	Joy	Sadness	Average	Anger	Fear	Joy	Sadness
Initial SVR	0.755	0.710	0.801	0.758	0.753	0.647	0.595	0.694	0.625	0.673
Fine-Tuned SVR	0.725	0.696	0.763	0.731	0.711	0.605	0.593	0.636	0.581	0.611

TABLE IX
COMPARISON WITH PREVIOUS RELATED STUDIES.

Methods	Pearson					Gold Pearson				
	Average	Anger	Fear	Joy	Sadness	Average	Anger	Fear	Joy	Sadness
Baseline	0.650	0.630	0.650	0.650	0.650	0.470	0.510	0.510	0.400	0.490
NUIG [12]	0.494	-0.047	0.680	0.717	0.625	0.390	0.003	0.567	0.566	0.426
SeerNet [11]	0.708	0.745	0.676	0.698	0.715	0.547	0.556	0.529	0.551	0.551
Prayas [9]	0.747	0.765	0.732	0.762	0.732	0.571	0.557	0.605	0.621	0.500
The research	0.755	0.710	0.801	0.758	0.753	0.647	0.595	0.694	0.625	0.673

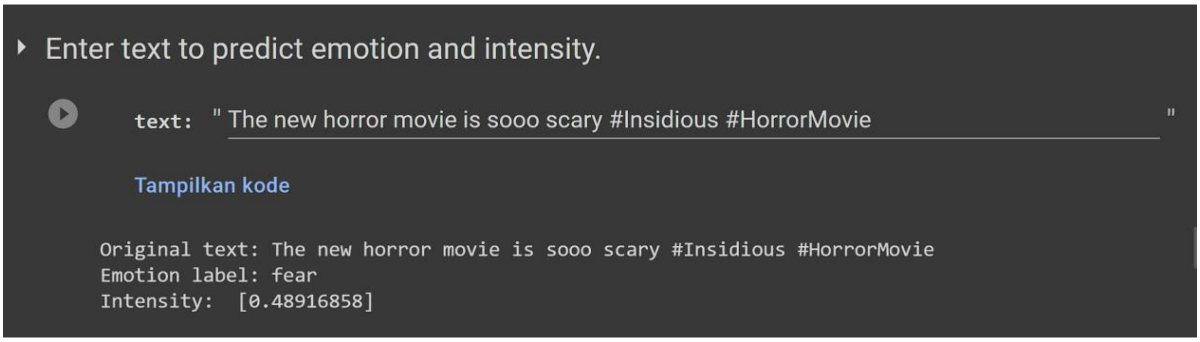

Fig. 2. New data prediction for joy (upper) and fear (lower).

the final models. The results show that SVR with features from TF-IDF performs better than others, with Ridge Regression placing second. Exploration of feature extraction parameters has been attempted and analyzed in the research. However, the change is relatively minuscule. The following study on the model hyperparameter has also been conducted through Grid Search, but the result obtained through this process is overfitting on the training dataset. It concludes that using a similar setup will unlikely obtain any significant improvement through parameter tuning. However, these results can be explored further by implementing different preprocessing techniques, which may improve or worsen the result depending on the fit of the method with the dataset. In addition, the research compares the performance of the final model with some previous related works. It shows a decent result comparable with previous research by performing better in fear, sadness, joy, and all the gold label value.

Nevertheless, it needs to be noted that the research does have several limitations. First, it lacks high-performance computing. Second, the research has been done on free GPU platform, Google Colab. Thus, the time taken to simulate the research may differ due to resource capability. For future research, collecting additional datasets and resources for these tasks will be highly valuable in improving the research due to the lack of resources in the current state. It can be done through unsupervised learning or manually annotating the curated data. Adding more modalities, such as speech and image, is another interesting approach to analyze.

References

[1] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang, and W. Zhang, “A systematic review on affective computing: Emotion models, databases, and recent advances,” *Information Fusion*, vol. 83, pp. 19–52, 2022.

[2] N. Mejbri, F. Essalmi, M. Jemni, and B. A. Alyoubi, “Trends in the use of affective computing in e-learning environments,” *Education and Information Technologies*, vol. 27, pp. 3867–3889, 2022.

[3] S. B. I. Badia, L. V. Quintero, M. S. Cameirao, A. Chirico, S. Triberti, P. Cipresso, and A. Gaggioli, “Toward emotionally adaptive virtual reality for mental health applications,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 5, pp. 1877–1887, 2018.

[4] M. Heo and K. J. Lee, “Chatbot as a new business communication tool: The case of naver talktalk,” *Business Communication Research and Practice*, vol. 1, no. 1, pp. 41–45, 2018.

[5] A. Kumar and A. Jaiswal, “Systematic literature review of sentiment analysis on twitter using soft computing techniques,” *Concurrency and Computation: Practice and Experience*, vol. 32, no. 1, 2020.

[6] S. Mohammad and F. Bravo-Marquez, “Emotion intensities in Tweets,” in *Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (SEM 2017)*. Vancouver, Canada: Association for Computational Linguistics, 2017, pp. 65–77.

[7] ——, “WASSA-2017 shared task on emotion intensity,” in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Copenhagen, Denmark: Association for Computational Linguistics, 2017, pp. 34–49.

[8] S. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, “SemEval-2018 task 1: Affect in Tweets,” in *Proceedings of the 12th International Workshop on Semantic Evaluation*. New Orleans, Louisiana: Association for Computational Linguistics, 2018, pp. 1–17.

[9] P. Goel, D. Kulshreshtha, P. Jain, and K. K. Shukla, “Prayas at EmoInt 2017: An ensemble of deep neural architectures for emotion intensity prediction in Tweets,” in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Copenhagen, Denmark: Association for Computational Linguistics, 2017, pp. 58–65.

[10] V. Duppada and S. Hiray, “Seernet at EmoInt-2017: Tweet emotion intensity estimator,” in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Copenhagen, Denmark: Association for Computational Linguistics, 2017, pp. 205–211.

[11] V. Andryushechkin, I. Wood, and J. O’Neill, “NUIG at EmoInt-2017: BiLSTM and SVR ensemble to detect emotion intensity,” in *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Copenhagen, Denmark: Association for Computational Linguistics, 2017, pp. 175–179.

[12] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, 2019, pp. 4171–4186.

[13] M. García, S. Maldonado, and C. Vairetti, “Efficient n-gram construction for text categorization using feature selection techniques,” *Intelligent Data Analysis*, vol. 25, no. 3, pp. 509–525, 2021.

[14] Z. Jiang, B. Gao, Y. He, Y. Han, P. Doyle, and Q. Zhu, “Text classification using novel term weighting scheme-based improved TF-IDF for Internet media reports,” *Mathematical Problems in Engineering*, vol. 2021, pp. 1–30, 2021.

[15] T. O. Olaleye, O. T. Arogundade, A. Abayomi-Alli, and A. K. Adesemowo, “An ensemble predictive analytics of COVID-19 infodemic tweets using bag of words,” in *Data science for COVID-19*. Elsevier, 2021, pp. 365–380.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau, “Scikit-learn: Machine learning in Python,” *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[17] I. Syarif, A. Prugel-Bennett, and G. Wills, “SVM parameter optimization using grid search and genetic algorithm to improve classification performance,” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 14, no. 4, pp. 1502–1509, 2016.

[18] M. A. Mohsin and A. Beltiukov, “Summarizing emotions from text using Plutchik’s wheel of emotions,” in *7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2019)*. Ufa, Russia: Atlantis Press, May 28–29, 2019, pp. 291–294.

Appendix

The Appendix can be seen in the next page.

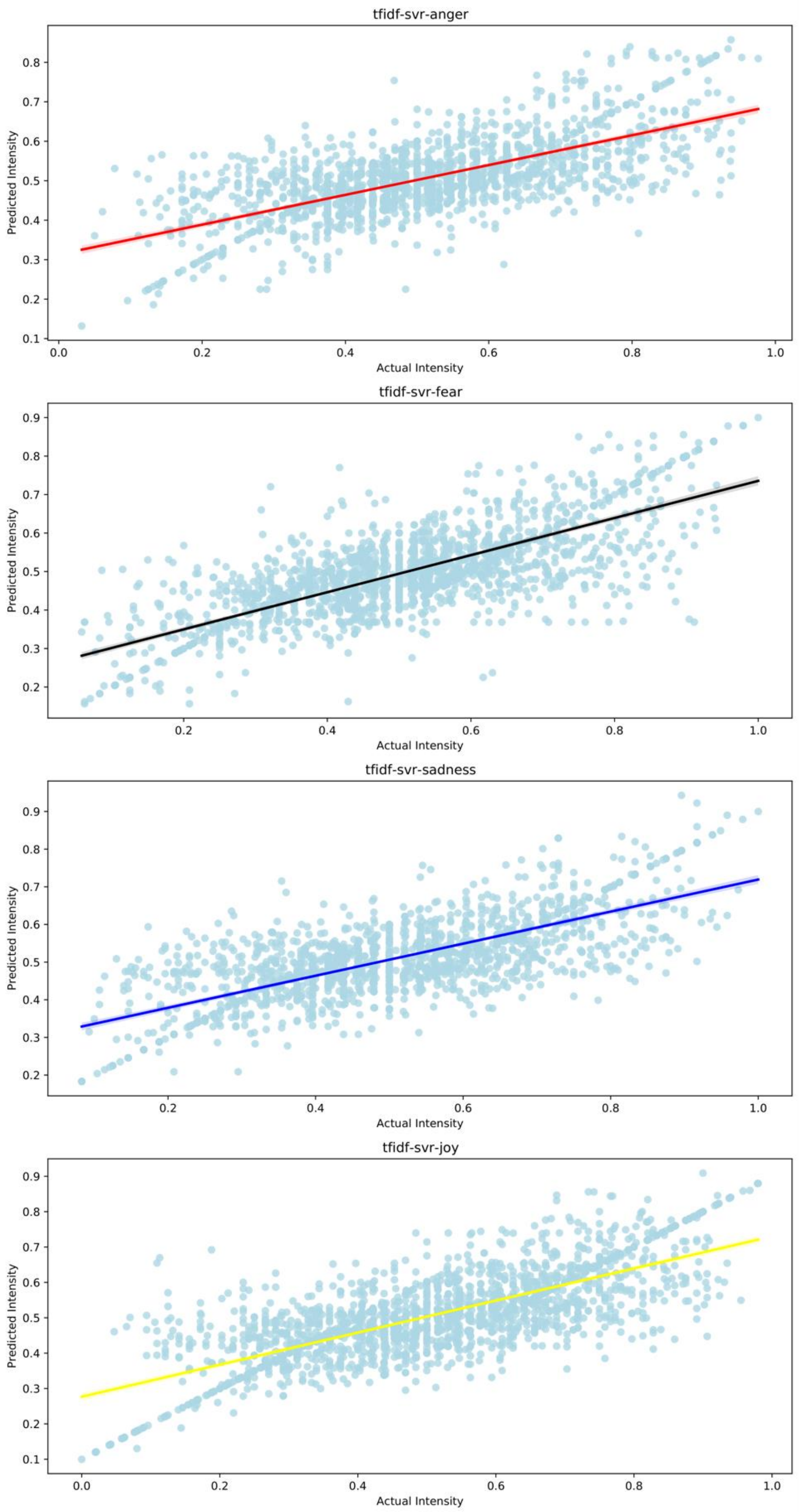

Fig. A1. Predicted intensity vs. actual intensity using TF-IDF SVR comparison.