

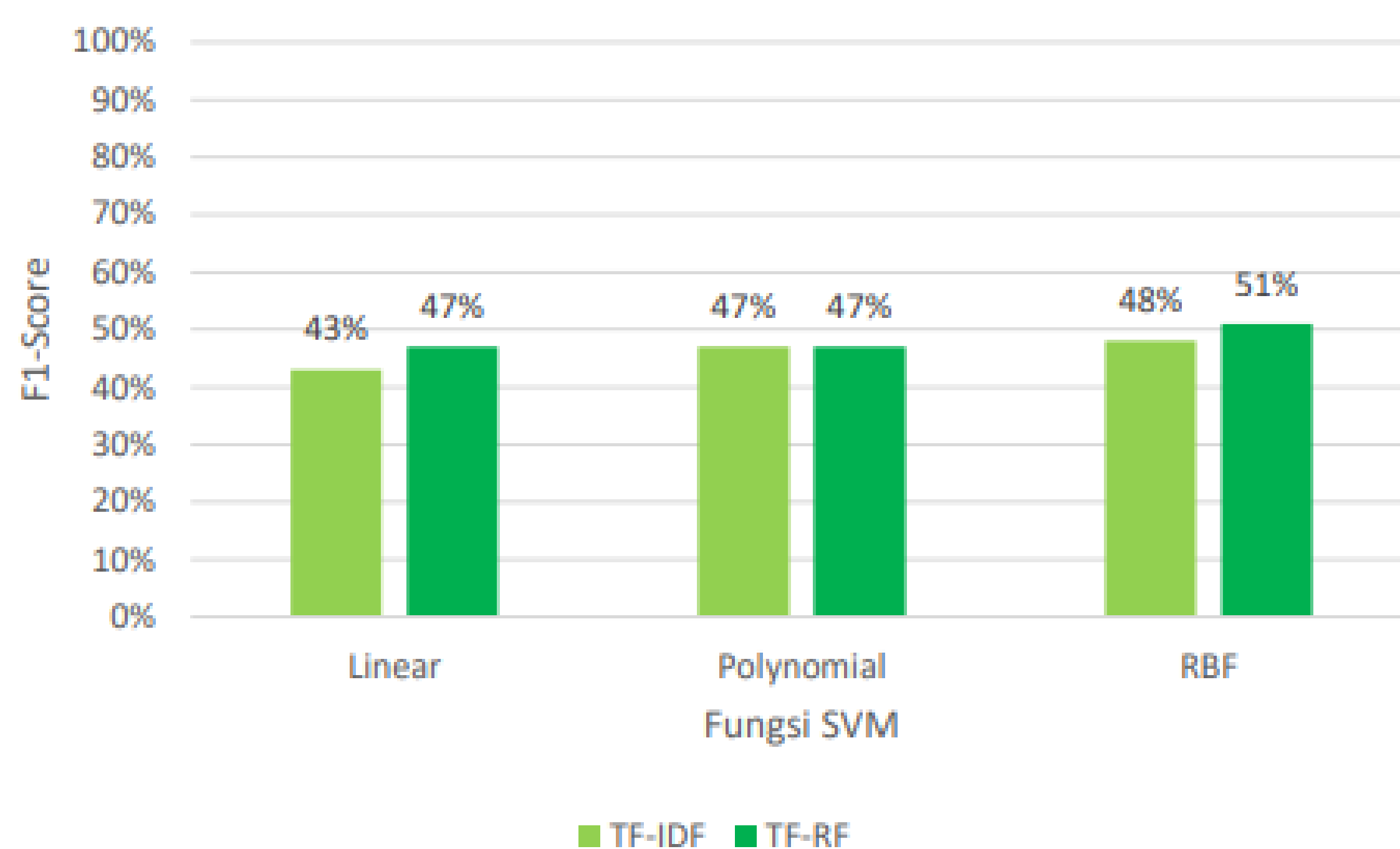
SENTIMENT ANALYSIS OF SEXUAL HARASSMENT TWEETS WITH COMPARISON OF TERM WEIGHTING METHODS USING SVM CLASSIFICATION AGAINST TAG PERMENDIKBUD30

MEIRA REYNITA PUTRI
1301164695

DR KEMAS MUSLIM LHAKSMANA, S.T., M.ISD
1382075

ABSTRACT

Twitter is one of the social media that is used as a means of expressing opinions and expressing themselves, both in channeling opinions and aspirations of the community as a form of democratic activity. One example is the ratification of the Minister of Education and Culture Regulation (Permendikbud) No. 30 of 2021 concerning the Prevention and Handling of Sexual Violence (PPKS) in Higher Education. The emergence of Tweets with the hashtag #permendikbud30 reaps the pros and cons among Twitter social media users. To process the Tweet information, sentiment analysis is carried out which serves to determine opinions or opinions about a product or event. In the process, Tweets are processed using data mining, namely classification. In determining the classification there are several stages that must be done, namely dataset, labeling, confusion matrix, weighting and accuracy results. Based on the system built, it will be seen from which weighting method has the highest accuracy value in sentiment analysis against #permendikbud30. Based on the test results, the highest F1-Score value for TF-RF with the SVM kernel rbf function is 51%.



TEST ANALYSIS

Based on the test results using the kernel function and gamma, C, and degree parameters, the highest F1-Score value is obtained in the linear kernel for TF-IDF = 43% and TF-RF = 47% with a value of $C = 1$. The F1-Score value is displayed because the amount of data is not balanced, then the polynomial kernel for each term weighting TF-IDF and TF-RF gets a value of 47% with a combination of gamma = 1 and degree = 1 and 2. While the rbf kernel for TF-IDF = 48% and TF-RF = 51% with a value of gamma = 1.0 and $C = 10$. Therefore, by using the rbf kernel, the F1-Score value increases by 4%. This is caused by the greater the value of the constant C, the greater the value of the F1-Score produced, while for the gamma the greater the value of the constant does not have much effect on the value of the F1-Score produced, so an optimum gamma value is needed to be able to produce the value of the produce a good F1-Score.

CONCLUSION

- Based on the test results, the F1-Score value for TF-IDF using the SVM kernel function experienced a significant increase.
- The F1-Score value for TF-RF is better than TF-IDF, reaching a value of 57% when using the rbf kernel function with a combination of gamma = 1.0 and $C = 10$ values.

SUGGESTION

In future research, it is possible to add datasets and or use other classification methods to increase the accuracy value so that the accuracy value obtained is higher.

REFERENCE

- N. Z. U. N. F. M. N. F. Kholiq Budiman, "Analysis of Sexsual Harassment Tweet Sentiment on Twitter in Indonesia using Naive Bayes Method through National Institute of Standard and Technology Digital Forensic Acquistition Approach," *Journal of Advances in Information Systems and Technology* 2 (2), vol. 21, pp. 21-30, 2020.
- X. L. Y. S. Emma Haddi, "The Role of Text Pre-processing in Sentiment Analysis," *ProcediaComputer Science* 17, pp. 26-32, 2013.