2024 International Conference on Information Technology Research and Innovation (ICITRI) | 979-8-3503-7621-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICITRI62858.2024.10699016

Comparison of Oversampling Techniques in Prediction Judicial Decisions of Divorce Trials in Family Courts

Aghnia Bella Dina
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
6025231072@student.its.ac.id

Riyanarto Sarno
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@ if.its.ac.id

Ratih Nur Esti Anggraini
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
ratih_nea@if.its.ac.id

Agus Tri Haryono
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7025231020@student.its.ac.id

Abdullah Faqih Septiyanto
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7025221022@if.its.ac.id

***Abstract*—This study delves into predicting decisions in family courts, explicitly addressing the challenges of imbalanced datasets in divorce trials. We evaluated four oversampling techniques to enhance machine learning model accuracy in these contexts—SMOTE, Random Over, GAN, and ADASYN. Historical Surabaya's Religious Court case records were preprocessed and divided into training and testing sets. Alongside TF-IDF vectorization, we incorporated Indo BERT to prepare the textual data for machine learning analysis. These oversampling methods were employed before model training to counteract class imbalance. An extensive grid search identified optimal hyperparameters, refining the performance of models like Support Vector Machines, Random Forest, and Naive Bayes. Models were assessed through accuracy, precision, recall, and F1-score metrics to determine the most effective oversampling and machine learning model combination. The most notable results were achieved with Random Forest paired with SMOTE, achieving a precision of 0.82%, a recall of 0.81%, an F1-Score of 0.80%, and an accuracy of 0.81%. This study highlights the significance of tailored oversampling strategies and advanced machine learning for accurate, data-driven decisions in family courts, emphasizing thorough data preparation and meticulous model tuning.**

Keywords—Divorce Prediction, SMOTE, Random Over, GAN, ADASYN, SVM, Legal Decision Making, Imbalanced Data

I. INTRODUCTION

Due to high divorce rates, an efficient family court system is crucial for accurate trial predictions, especially in child custody cases, to protect children's rights and enhance court efficiency [1]**.** Technological advancements have led to more efficient legal and governmental systems [2]. However, an imbalance in divorce datasets may reduce prediction accuracy, requiring balancing techniques to ensure unbiased decision-making [3]. Class imbalance skews predictions, necessitating data balancing for improved accuracy [4], [5].

Machine learning can analyze legal cases, but unbalanced data leads to limited and poor predictions of trial outcomes [6–8]. Various methods have been created by researchers, and one of them is random oversampling, Deep Learning models with diverse configurations, and optimal classifiers, improving prediction accuracy [8], [9].

Various data balancing techniques are employed to enhance the accuracy of divorce predictions in family courts. The method referred to Synthetic Minority Oversampling Technique (SMOTE) and Random Over Sampling (ROS) [10], [11]. The ADAptive SYNthetic (ADASYN) technique uses k-nearest neighbors to generate minority data samples adaptively, updating the distribution without assuming data uniformity. Generative Adversarial Networks (GANs) create synthetic data through deep learning.

This study advances family law by analyzing divorce decisions in religious courts using four oversampling techniques (SMOTE, Random Over, GAN, and ADASYN) on an imbalanced dataset from the Religious Court of Surabaya. It enhances the accuracy of models in machine learning like SVM, Random Forest, and Naive Bayes. It demonstrates how tailored oversampling and advanced machine learning can lead to fairer, data-driven decision-making in family courts.

This research aims to enhance information technology and legal practice by applying machine learning to legal data, offering insights for legal system stakeholders. It advances technical knowledge in legal contexts and improves decision-making in family courts, bridging the gap between technology and law. A predictive system can enhance the transparency and fairness of the judicial process [12], [13]. However, it's crucial to clearly define the experiment's purpose and ensure the system's meaningful contribution to the legal field[14].

The paper's structure is as follows: Section II covers relevant literature, Section III details methodology, Section IV presents results and discussion, and Section V summarizes findings, implications, and future research directions.

II. LITERATURE STUDY

This literature review comprehensively examines a spectrum of studies, each contributing to the burgeoning

fields of data balancing and machine learning, with a specific focus on applications in legal and psychological analysis and natural language processing.

At the center of challenges in processing natural language, the work [15] sheds light on the intricacies of feature extraction in text document classification, employing the TF-IDF technique. Their research delves into the complexities inherent in understanding and interpreting the multifaceted meanings and expressions found in documents, highlighting the significant role of feature extraction in accurately labelling and classifying text.

In social media and information dissemination, the capabilities of IndoBERT, a model fine-tuned explicitly for the Indonesian language, underline its effectiveness in identifying and curtailing the spread of disinformation. This research uniquely combines IndoBERT with NLTK (Natural Language Tool Kit) Tokenizer and BERT (Bidirectional Encoder Representation from Transformer) AutoTokenizer, achieving high accuracy rates and offering novel insights into combating misinformation in digital communication [16].

Shifting the focus to data balancing techniques, a notable study on the Synthetic Minority Oversampling Technique (SMOTE) showcases its utility in enhancing machine learning models' accuracy. Specifically, the study highlights how the RVVC model, when applied to balanced datasets using SMOTE, outperforms other state-of-the-art approaches, especially when handling datasets with varied feature vector sizes. This research underscores SMOTE's potential in refining the performance of predictive models across diverse scenarios [17].

In addressing the issue of class imbalance in hate speech detection, the research conducted by [18] is particularly enlightening. It explores four resampling techniques—ROS, SMOTE, ADASYN, and RUS—in combination with fundamental machine learning classifiers like Support Vector Machine, Logistic Regression, and Naive Bayes. The study concludes that oversampling methods, especially ROS, significantly enhance these classifiers' accuracy and overall performance, offering a pragmatic remedy for addressing the difficulties presented by imbalanced datasets in hate speech detection.

Another research investigated the impact of data balancing methods on the classification of legal documents decision on Consumer Rights. In addition to common evaluation metric such as accuracy and F1-score, this study also evaluated the results using G-mean and kappa to avoid misleading conclusions. The results showed that SMOTE and Borderline SMOTE as the most effective data balancing methods for predicting the text legal documents [19].

Generative Adversarial Networks (GANs), a revolutionary approach in machine learning, are the focus of a study by [20] to address class imbalance in classification models. By creating realistic samples of underrepresented classes, their research highlights GANs' potential in generating balanced datasets, particularly in credit card fraud detection.

The increasing trend of divorces and its societal impact form the backdrop of studies by [21] and [22]. Machine learning, including the Perceptron model, was used in [21] to predict divorce rates, aiding counsellors with high accuracy. Meanwhile, psychological lens and machine learning was applied by [22] to forecast divorces from expert-identified patterns, offering insights for psychologists and family counsellors.

III. METHODOLOGY

In this section, as illustrated in Figure 1, In our study, the methodology includes dataset pre-processing, data oversampling, machine learning algorithm selection, model training, and prediction using the trained model.

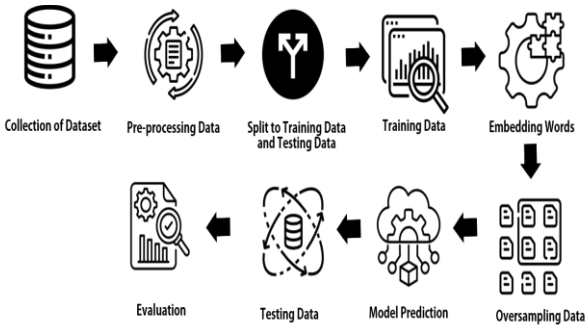

Fig. 1. Architecture for predicting court decision

In the following segment, we provide an elaborate account of our methodology, which is comprehensively illustrated in Figure 2. This figure serves as a visual guide, elucidating each step of our process, from the initial data collection to the final stages of model evaluation. It provides a clear and concise roadmap of our research approach, ensuring readers can easily understand and follow the systematic procedures we employed in this study. Additionally, the figure highlights the integration of diverse data sources and advanced analytical techniques, which are crucial for the robustness and reliability of our findings.

A. Collection of Dataset

In our analysis, we employed a dataset from the Surabaya Religious Court, concentrating on four pivotal attributes: 'posita,' 'status_putusan' (status of the decision), 'amar_putusan' (verdict announcement), and 'label.' The 'posita' is the case's introductory statement, providing context and forming the basis for prediction. 'Amar_putusan' encapsulates the court's final decision, while 'status_putusan' indicates if a divorce is granted. These attributes inform the 'label', the prediction target in our model. Table 1 below outlines these features and their importance in the predictive model. This dataset, comprising 5,919 data, is categorized into three labels: Label 0 represents cases where the divorce petition was denied, Label 1 corresponds to cases where the divorce was granted with child custody rights, and Label 2 indicates divorces granted without child custody rights. The dataset exhibits an imbalanced composition, with 883 cases under Label 0, 1,625 cases under Label 1, and 3,411 cases under Label 2. In our study, we tackle a class imbalance ratio of roughly 1:2 in our dataset, mirroring situations often encountered in data analytics. Such an imbalance is in line with cases observed in well-known datasets [23].

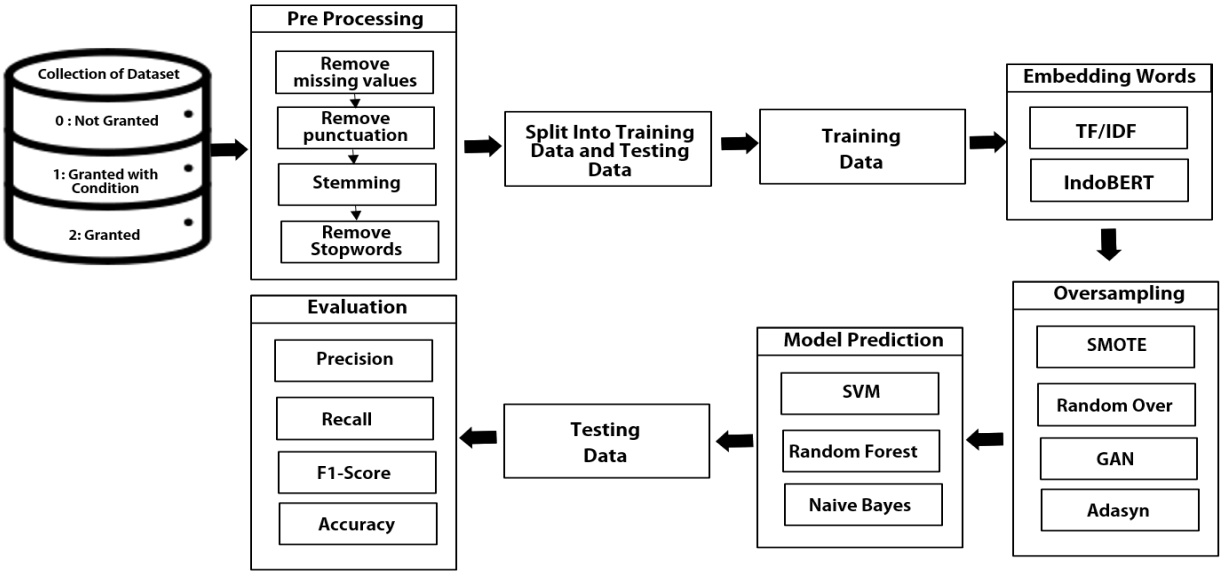

Fig. 2. Detailed overview of predicting court decisions

TABLE I. DESCRIPTION OF DATASET ATTRIBUTES

Attribute Name	*Description*	*Data Type*
Posita	Introductory statement of the case, containing facts and grounds.	Text
amar_putusan /announced the verdict	Contains the court's verdict or decision on the case.	Text
status_putusan/'announced the verdict	Records the status of the divorce case decision, indicating whether the petition is accepted or denied.	Categorical
label	Categories case outcomes based on 'amar_putusan,' with specific categories for different verdict types.	Categorical

Procedure:
Start
Step #01: Take a posita from the dataset in the previous stage
Step #02: Use the Python Natural Language Toolkit (NLTK) to
doing pre-processing text
Step #03: Do Remove Missing Values → Remove punctuation → Stemming → Remove Stop Words
Step #04: Save the result for the next stage.
End

Fig. 3. Pseudocode outlining the text pre-processing steps

B. Data pre-processing

Figure 3 displays the pseudocode for our standard text preprocessing procedure. In this process, we also eliminate lines with empty posita. Those steps in data preprocessing were needed because some data has missing values, where these conditions can affect the result. In addition, stemming and stop word removal were needed so the data is ready to be processed which is word embedding.

C. Word Embedding

Subsequently, we undertook several steps to train the dataset after the text preprocessing process

a) TF-IDF (Term Frequency-Inverse Document Frequency: The TF-IDF an abbreviation Term Frequency-Inverse Document Frequency, metric assigns a significance level to each word in a document, escalating with the word's occurrence within that document but offset by its prevalence across all records in the dataset [24].

b) IndoBERT: IndoBERT Word Embedding, though not directly referenced in the search findings, is a pre-trained model explicitly designed for the Indonesian language. It is a variant of the widely recognized BERT model formally recognized as Bidirectional Encoder Representation from Transformers. The training of the IndoBERT model is conducted on the Indo4B dataset, encompassing a comprehensive collection of Indonesian news text s[25], [26]. This training has equipped IndoBERT to be adept at various Natural Language Processing (NLP) tasks [27].

D. Oversampling

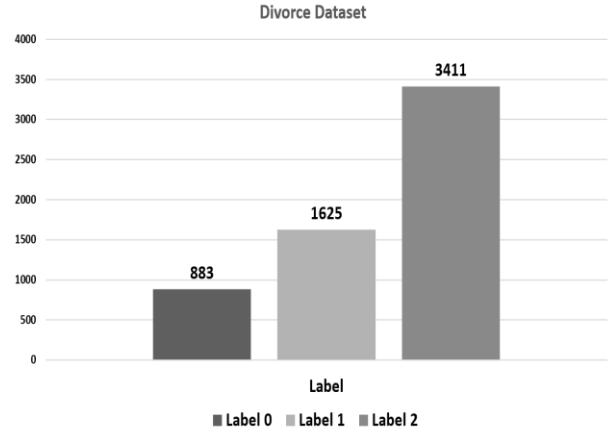

Fig. 4 Imbalanced Dataset

After vectorizing TF-IDF and IndoBert, we applied random oversampling to address this imbalance problem. Figure 4 above illustrates that the composition of the dataset needs to be more balanced. In this study, we applied four varied oversampling techniques, which are listed below:

a) SMOTE: SMOTE (Synthetic Minority Over-sampling Technique) is used for augment underrepresented classes in machine learning datasets to mitigate class imbalances. It produces synthetic instances that are excessively similar to the original samples of the minority class, thereby not adequately expanding the decision boundary [27].

b) Random Over: This technique, while simple, effectively addresses the class imbalance in datasets, a common issue in machine learning. By equalizing class sizes, ROS enhances the model's ability to learn from underrepresented data, leading to more accurate and generalized predictions [28].

c) ADASYN: ADASYN, an advanced oversampling technique, produces synthetic data for minority class samples that are more difficult to learn, reducing class imbalance bias and adjusting the classification boundary. Its effectiveness is proven in simulations evaluated across five metrics [30].

*d) GAN: T*he Generative Adversarial Network (GAN) features two components: a generative block creating synthetic data resembling actual samples and a discriminative block distinguishing natural and synthetic data. This structure allows GANs to enhance artificial data quality through ongoing interaction between the two blocks [31].

E. Machine Learning

a) Split Training Data and Testing Data: In this study section, the dataset is segmented into training and testing sets, 80% allocating for training and 20% for testing. The training data undergoes vectorization, Transforming textual information into a numerical representation that is understandable and usable by machine learning algorithms.

b) SVM: SVM, which stands for Support Vector Machine, is an algorithm used in supervised learning that categorizes data points into classes. It maximizes the margin between data points of different types to create an effective decision boundary [20].

c) Random Forest: A Random Forest is made up of an ensemble of decision trees. Each tree has leaf nodes representing class labels and intermediate nodes containing independent variables, including binary choices. The ultimate decision for classification is determined by the majority vote from all the trees, thereby increasing the accuracy and reliability of the prediction.

d) Naive Bayes: Naive Bayes is a probabilistic classifier that uses Bayes' Theorem to make predictions [12]. Each feature is assumed to be independent and equally influential in this model. Naive Bayes is known for its speed, often outperforming other classifiers regarding computational efficiency [20].

e) Grid Search Cross Validation: In machine learning, grid search cross-validation is used for fine-tuning hyperparameters, the non-learned parameters within models. It systematically explores a range of hyperparameter values, using cross-validation to evaluate each combination's performance. This process divides the data into several subsets, using each in turn for validation and the rest for training.

K-Fold Cross Validation: Cross-validation (CV) plays crucial for assessing models' generalization and predictive strength in data mining and machine learning. It entails splitting the dataset into training and testing subsets, utilizing one for model training and the other for performance evaluation, which is essential in model or method selection [30].

F. Evaluation

In this research, we assessed our machine learning models using four essential metrics: precision, recall, F1-score, and accuracy. Precision shows the fraction of accurate positive forecasts among the total of optimistic predictions, underscoring the model's accuracy in effectively pinpointing relevant examples. Recall, alternatively called sensitivity, gauges the model's proficiency in detecting positive cases. The F1-score, combines the harmonic average of precision and recall, serves as a balanced metric, considering both precision and memory, and is particularly beneficial when balancing these two measures is essential. Lastly, accuracy indicates the overall effectiveness of the model by measuring the fraction of correct predictions (encompassing true positives and true negatives) among the total instances.

IV. Results and Discussions

A. Result

TABLE II. Table Performance of TF-IDF Word Embedding Across Various Oversampling Methods and Techniques

Metode	TF-IDF			
	Precision	***Recall***	***F1-Score***	***Accuracy***
SVM	0.79	0.78	0.77	0.78
SVM-SMOTE	0.77	0.77	0.76	0.77
SVM - RANDOMOVER	0.77	0.78	0.78	0.77
SVM - ADASYN	0.77	0.77	0.77	0.77
SVM - GAN	0.79	0.78	0.77	0.78
Random Forest	0.81	0.79	0.77	0.79
Random Forest - SMOTE	**0.82**	**0.81**	**0.80**	**0.81**
Random Forest - RANDOMOVER	0.81	0.80	0.79	0.80
Random Forest - ADASYN	0.77	0.77	0.76	0.77
Random Forest - GAN	0.82	0.80	0.78	0.80
Naive Bayes	0.73	0.74	0.71	0.73
Naive Bayes - SMOTE	0.66	0.64	0.65	0.64
Naive Bayes-RANDOMOVER	0.67	0.65	0.66	0.65
Naive Bayes - ADASYN	0.68	0.65	0.66	0.65
Naive Bayes - GAN	0.64	0.68	0.60	0.68

The assessment of 28 simulations with various oversampling techniques revealed key findings. The Random Forest-SVM method using TF-IDF embedding achieved the highest accuracy, especially in religious divorce cases, highlighting oversampling's effectiveness. Overall, oversampling consistently improved model accuracy. However, Naive Bayes-Adasyn with IndoBERT embedding had the lowest performance, indicating a need for optimization.

Specifically, the highest accuracy was achieved in the Random Forest - SVM method using TF-IDF and SVM using IndoBERT, showing a significant improvement compared to the baseline models. On average, simulations with TF-IDF and the Naive Bayes method demonstrated balanced performance improvements across various scenarios. Although Naive Bayes-Adasyn using IndoBERT had the lowest performance, it provides insights into the limitations of the current modeling approach.

TABLE III. TABLE PERFORMANCE OF INDOBERT WORD EMBEDDING ACROSS VARIOUS OVERSAMPLING METHODS AND TECHNIQUES

Metode	IndoBERT			
	Precision	***Recall***	***F1-Score***	***Accuracy***
SVM	**0.82**	**0.81**	**0.80**	**0.81**
SVM-SMOTE	0.77	0.76	0.74	0.76
SVM - RANDOMOVER	0.78	0.77	0.75	0.77
SVM - ADASYN	0.77	0.76	0.74	0.76
SVM - GAN	0.82	0.80	0.78	0.80
Random Forest	0.78	0.76	0.75	0.76
Random Forest - SMOTE	0.78	0.76	0.75	0.76
Random Forest - RANDOMOVER	0.76	0.76	0.75	0.76
Random Forest - ADASYN	0.75	0.75	0.75	0.75
Random Forest - GAN	0.77	0.76	0.75	0.76
Naive Bayes	0.65	0.63	0.63	0.63
Naive Bayes - SMOTE	0.68	0.64	0.64	0.64
Naive Bayes-RANDOMOVER	0.65	0.62	0.62	0.62
Naive Bayes - ADASYN	0.58	0.54	0.55	0.54
Naive Bayes - GAN	0.65	0.63	0.63	0.63

To gain deeper insights and a thorough comprehension of the performance of each simulation, refer to Table II for TF-IDF simulations and Table III for IndoBERT simulations.

B. Discussion

Our study's findings reveal valuable insights into oversampling techniques in legal analytics. Initially, the highest-performing models were Random Forest with TF-IDF word embedding and SVM with IndoBERT embedding, displaying predictive solid capabilities. However, after applying SMOTE oversampling, a notable performance boost occurred in the Random Forest model with TF-IDF embedding, emphasizing the significance of data balance in legal predictive models. This finding is also in line with the previous research [17, 19] stated that SMOTE outperformed other oversampling methods.

For IndoBERT embedding, oversampling didn't enhance any model's performance compared to SVM alone, indicating that model efficacy varies with embedding techniques and highlighting the importance of tailored model selection and data preparation—moreover, the integration of oversampling techniques led to shifts in model superiority. While Random Forest initially excelled, using SMOTE made it the most effective approach, highlighting the dynamic nature of predictive modeling and the impact of data preparation techniques like oversampling.

The comparison between IndoBERT embedding and TF-IDF feature extraction yielded valuable insights. Although IndoBERT initially showed promise, every model could only match SVM with oversampling after applying oversampling. This raises questions about the compatibility and effectiveness of various embedding and oversampling combinations, emphasizing the necessity for further exploration and optimization in legal analytics.

V. CONCLUSION AND FUTURE WORK

This research signifies a significant advancement in predicting divorce outcomes within family courts, particularly in addressing challenges related to imbalanced datasets. The effective combination of advanced text embedding techniques, notably SMOTE, with sophisticated machine learning models, particularly Random Forest, demonstrated high accuracy and predictive reliability, highlighting the power of merging specialized text processing and machine learning in legal analytics.

In future work endeavours, we plan to refine our approach by adopting word embedding techniques such as word2Vec, GloVe or fastText and compare it with traditional embedding method that is used in this research to achieve a better performance. Moreover, expanding it into more diverse legal cases will improve the model applicability and increase the usefulness on real cases.

ACKOWLEGEMENT

This research was funded by Institut Teknologi Sepuluh Nopember (ITS) under Penelitian Dana Departemen Program, Penelitian Kerjasama Antar Perguruan Tinggi, and under the Project Scheme of the Publication Writing and Intellectual Property Rights (IPR) Incentive Program 2024.

REFERENCES

[1] C. S. Stover, “Commentary: Factors Predicting Family Court Decisions in High-Conflict Divorce,” 2013.

[2] M. Pise and M. Rameshrao Pise, “Significance of Technology in Family Courts: An Analysis. Family & Adoption Law Significance of Technology in Family Courts: An Analysis,” vol. 4, no. 2, p. 2021, 2021, doi: 10.37591/JFAL.

[3] K. W. Brown and P. C. Cozby, *Research methods in human development*. Mayfield Pub, 1999.

[4] SVS College of Engineering and Institute of Electrical and Electronics Engineers, *Proceedings of the 2018 International Conference on Current Trends towards Converging Technologies : 01-03, March 2018*.

[5] Z. Qureshi *et al.*, “Efficient Prediction of Missed Clinical Appointment Using Machine Learning,” *Comput Math Methods Med*, vol. 2021, pp. 1–10, Oct. 2021, doi: 10.1155/2021/2376391.

[6] C. Steging, S. Renooij, and B. Verheij, "Taking the Law More Seriously by Investigating Design Choices in Machine Learning Prediction Research," 2023. [Online]. Available: http://ceur-ws.org.

[7] M. F. Sert, E. Yıldırım, and İ. Haşlak, "Using Artificial Intelligence to Predict Decisions of the Turkish Constitutional Court," *Soc Sci Comput Rev*, vol. 40, no. 6, pp. 1416–1435, Dec. 2022, doi: 10.1177/08944393211010398.

[8] D. Alghazzawi, O. Bamasag, A. Albeshri, I. Sana, H. Ullah, and M. Z. Asghar, "Efficient Prediction of Court Judgments Using an LSTM+CNN Neural Network Model with an Optimal Feature Set," *Mathematics*, vol. 10, no. 5, p. 683, Feb. 2022, doi: 10.3390/math10050683.

[9] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and Y. D. Al-Otaibi, "A hybrid CNN + BILSTM deep learning-based DSS for efficient prediction of judicial case decisions," *Expert Syst Appl*, vol. 209, p. 118318, Dec. 2022, doi: 10.1016/j.eswa.2022.118318.

[10] G. A. Pradipta, R. Wardoyo, A. Musdholifah, I. N. H. Sanjaya, and M. Ismail, "SMOTE for Handling Imbalanced Data Problem : A Review," in *2021 Sixth International Conference on Informatics and Computing (ICIC)*, IEEE, Nov. 2021, pp. 1–8. doi: 10.1109/ICIC54025.2021.9632912.

[11] F. Kamalov, H.-H. Leung, and A. K. Cherukuri, "Keep it simple: random oversampling for imbalanced data," in *2023 Advances in Science and Engineering Technology International Conferences (ASET)*, IEEE, Feb. 2023, pp. 1–4. doi: 10.1109/ASET56582.2023.10180891.

[12] J. Dressel and H. Farid, "The accuracy, fairness, and limits of predicting recidivism," *Sci Adv*, vol. 4, no. 1, Jan. 2018, doi: 10.1126/sciadv.aao5580.

[13] J. Dressel and H. Farid, "The Dangers of Risk Prediction in the Criminal Justice System," *MIT Case Studies in Social and Ethical Responsibilities of Computing*, Feb. 2021, doi: 10.21428/2c646de5.f5896f9f.

[14] M. Medvedeva, M. Wieling, and M. Vols, "Rethinking the field of automatic prediction of court decisions," *Artif Intell Law (Dordr)*, vol. 31, no. 1, pp. 195–212, Mar. 2023, doi: 10.1007/s10506-021-09306-3.

[15] Y. Setiawan, D. Gunawan, and R. Efendi, "Feature Extraction TF-IDF to Perform Cyberbullying Text Classification: A Literature Review and Future Research Direction," in *2022 International Conference on Information Technology Systems and Innovation (ICITSI)*, IEEE, Nov. 2022, pp. 283–288. doi: 10.1109/ICITSI56531.2022.9970942.

[16] A. B. Y. A. Putra, Y. Sibaroni, and A. F. Ihsan, "Disinformation Detection on 2024 Indonesia Presidential Election using IndoBERT," in *2023 International Conference on Data Science and Its Applications (ICoDSA)*, IEEE, Aug. 2023, pp. 350–355. doi: 10.1109/ICoDSA58501.2023.10277572.

[17] V. Rupapara, F. Rustam, H. F. Shahzad, A. Mehmood, I. Ashraf, and G. S. Choi, "Impact of SMOTE on Imbalanced Text Features for Toxic Comments Classification Using RVVC Model," *IEEE Access*, vol. 9, pp. 78621–78634, 2021, doi: 10.1109/ACCESS.2021.3083638.

[18] H. Rathpisey and T. B. Adji, "Handling Imbalance Issue in Hate Speech Classification using Sampling-based Methods," in *2019 5th International Conference on Science in Information Technology (ICSITech)*, IEEE, Oct. 2019, pp. 193–198. doi: 10.1109/ICSITech46713.2019.8987500.

[19] D. L. Freire, A.M. de Almeida, M. de S. Dias, A. Rivolli, F.S. Pereira, G.A. de Godoi, & A.C. de Carvalho, "Exploratory Study of Data Sampling Methods for Imbalanced Legal Text Classification", in *International Conference on Hybrid Artificial Intelligence Systems*, Aug. 2023, pp. 108-120, Cham: Springer Nature Switzerland. doi: https://doi.org/10.1007/978-3-031-40725-3_10.

[20] H. Bhagwani, S. Agarwal, A. Kodipalli, and R. J. Martis, "Targeting class imbalance problem using GAN," in *2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies, and Optimization Techniques (ICEECCOT)*, IEEE, Dec. 2021, pp. 318–322. doi: 10.1109/ICEECCOT52851.2021.9708011.

[21] A. Sharma, A. S. Chudhey, and M. Singh, "Divorce case prediction using Machine learning algorithms," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, IEEE, Mar. 2021, pp. 214–219. doi: 10.1109/ICAIS50930.2021.9395860.

[22] C. Oswald, S. Baranwal, S. M. S. S. Narayanan, and A. Bhattacharya, "Divorce Astrologer: Machine Learning based Divorce Prediction of Married Couples," in *2022 IEEE 19th India Council International Conference (INDICON)*, IEEE, Nov. 2022, pp. 1–6. doi: 10.1109/INDICON56171.2022.10040167.

[23] A. Gosain and S. Sardana, "Handling class imbalance problem using oversampling techniques: A review," in *2017 International Conference on Advances in Computing, Communications, and Informatics (ICACCI)*, IEEE, Sep. 2017, pp. 79–85. doi: 10.1109/ICACCI.2017.8125820.

[24] L. Pradhan, C. Zhang, S. Bethard, and X. Chen, "Embedding User Behavioral Aspect in TF-IDF Like Representation," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, IEEE, Apr. 2018, pp. 262–267. doi: 10.1109/MIPR.2018.00061.

[25] B. V. Kartika, M. J. Alfredo, and G. P. Kusuma, "Fine-Tuned IndoBERT Based Model and Data Augmentation for Indonesian Language Paraphrase Identification," *Revue d'Intelligence Artificielle*, vol. 37, no. 3, pp. 733–743, Jun. 2023, doi: 10.18280/ria.370322.

[26] S. O. Khairunnisa, Z. Chen, and M. Komachi, "Dataset Enhancement and Multilingual Transfer for Named Entity Recognition in the Indonesian Language," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 6, pp. 1–21, Jun. 2023, doi: 10.1145/3592854.

[27] R. Blagus and L. Lusa, "SMOTE for high-dimensional class-imbalanced data," *BMC Bioinformatics*, vol. 14, no. 1, p. 106, Dec. 2013, doi: 10.1186/1471-2105-14-106.

[28] N. S. Rahmi, N. W. S. Wardhani, M. B. Mitakda, R. S. Fauztina, and I. Salsabila, "SMOTE Classification and Random Oversampling Naive Bayes in Imbalanced Data : (Case Study of Early Detection of Cervical Cancer in Indonesia)," in *2022 IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA)*, IEEE, Nov. 2022, pp. 1–6. doi: 10.1109/ICITDA55840.2022.9971421.

[29] Haibo He, Yang Bai, E. A. Garcia, and Shutao Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, IEEE, Jun. 2008, pp. 1322–1328. doi: 10.1109/IJCNN.2008.4633969.

[30] A. Salazar, L. Vergara, and G. Safont, "New applications of an oversampling method based on generative adversarial networks," in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, IEEE, Dec. 2020, pp. 1699–1701. doi: 10.1109/CSCI51800.2020.00314.

[31] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," in *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, IEEE, Feb. 2016, pp. 78–83. doi: 10.1109/IACC.2016.25.