

Design of Personalization Exam Classification Model Based on Imbalanced Class

Sri Anita
Faculty Technology
Pertwi University
Bekasi, Indonesia
sri.anita@pertiwi.ac.id

Arif Rachmat
Faculty ScienceTech
Datokarama State Islamic
University
Palu, Indonesia
a@uindatokarama.ac.id

Sunu Aditya Mahadany
Technology Division
PT Integra Solusi Teknotama
Bandung, Indonesia
monstm@gmail.com

Abstract— Currently, personalized learning has become a necessity in the learning process. Research and implementation of personalization in the planning phases and implementation phases of learning have been extensively studied. However, that research has not yet reached the application stage in the learning assessment phase. Providing homogeneous examination material to each student has not considered the characteristics of learners. Even though the achievements from the assessment phase will provide a measure of the quality of the learning process as a whole. This research has analyzed the individual characteristics model, which is derived as a benchmark for identification of the information and characteristics of the test material, which is then formulated into a classification model based on supervised learning. This study identified text dataset questions and labeled unbalanced multi-classes. This presents a challenge to carry out experiments to find the most optimal data training strategy, the results provide optimal strategy combination results Logic: ENS, Verbal: ENS, Visual: CW+RES+ENS, Natural: CW+RES+ENS. Accuracy measurement results Logic (SVM): 0.85, Verbal (LR) 0.87, Visual (LR) : 0.93, Natural (NN) 0.93.

Keywords— personalization learning, personalization exam, imbalanced-class, multiple intelligence

I. INTRODUCTION

Personalization of learning is a learning strategy [1], [2], [3], [4], curriculum and learning environment designed to meet the needs and learning styles of each student [5], [6], [7]. According to [8], there are three aspects that can be adopted to differentiate the learning process for each individual.

- a Personalization, which is instruction that is tailored to your needs, preferences, and interests and also relates to diverse goals and content.
- b Differentiation, that is, instruction that is adapted to different ways of learning and methods.
- c Individualization, that is, instruction tailored to individual needs, with the same goal but with an adjusted learning speed.

Personalization in the context of information can be interpreted as negotiation between materials and information or user profiles. Therefore, both structure and information about the user are [9]–[12] needed before personalization can occur. In learning there are different points of view between personalization, differentiation and individualization of learning. The individualization of learning is realized as instruction that is tailored to the needs of each one, but has the same learning objectives, taking into account the tempo and needs of the material that do not have to be uniform.

II. PREVIOUS RESEARCH

This research deepens the development of a learning personalization system, especially in the stage of evaluating learning based on material classification with supervised learning techniques. Table 1. summarizes previous research related to exam personalization, formative tests, and assessments that take a solution approach in the field of information technology.

- 1) E-Assessment meets Personalization: Saul 2013 [13], model: Adaptive testing (Weiss dan Kingsbury 1984), Adaptive question (Tzanavari, Retalis, dan Pastellis 2004), methods: Rule-based model.
- 2) Intelligent Assessment and Content Personalization in Adaptive Educational Systems: McCusker et al. 2013 [14], Learning Styles Helper's Guide (Mumford 2006), methods: K-Nearest network.
- 3) Personalization of Tests for Formative Self-Assessment: Benchoff, González, dan Huapaya 2018 [15], Index of Learning Styles (ILS) (Soloman, Carolina, dan Felder 1999), methods: ILS-based assessment: Soloman et al., 2005, and previous knowledge (PK) survey.
- 4) An Ontology for Personalization in Serious Games for Assessment : Said, Cheniti-Belcadhi, dan El Khayat 2019 [6], Player's gameplay style (Nacke, Bateman, dan Mandryk 2014), methods: Ontology model
- 5) Classification of Exam Materials Based on Compound Intelligence Types, Multiple Intelligences (Gardner 1983), methods: Supervised learning.

III. RESEARCH METHODOLOGY

This study adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework [16], [17] as a work procedure, explanation of each stage in the development of the model in this study is as follows.

A. Business Understanding

At this stage, the initiation of the purpose of model development is to develop a model that can classify the written test manuscript. Then detailed the boundaries that will be set, in this case it will involve four types of classification, namely verbal, visual, logical and natural. Setting problem boundaries on the purpose of classifying provides focus in handling data at the next stage.

B. Data Understanding

The Two types of intelligencen aspects of knowledge were chosen, namely verbal intelligence and logic, and two types in the aspect of skills, namely visual and natural intelligence.

- 1) Verbal intelligence can take advantage of stories that have characters in line with this type of intelligence [18]. According to Abidia by Raharjo 2009, a story is a question that is presented in the form of a short story. The story revealed can be a matter of daily life or any other problem. The weight of the problem revealed will affect the short length of the story. Questions that can be used to determine students' abilities in mathematics can take the form of stories and questions not counting questions by Haji 1994.
- 2) Logical intelligence puts forward questions that hone the logic of thinking. [19] In basic education materials, students are given the ability to compare, define variables, draw opposite conclusions, material that displays mathematical symbols, as well as material dominated by calculations by Karso 2002 [19]. In

addition, classifying or sorting objects based on certain criteria is also an advantage in this intelligence.

- 3) Visual intelligence uses images and videos as the main objects in the question material. Visualization of form is more helpful for learners in understanding the material given. Narratives that invite students to imagine shapes, patterns, colors, textures are also things that stand out.
- 4) Natural intelligence uses natural objects as material to form question material. Grouping animals and plants is often found in the material. Illustrations of nature and natural phenomena also have an influence on the understanding of matter.

Table 1 are present the four types of intelligence are further analyzed based on the characteristics of learning content [20], [21], by identifying the appropriate adaptations to the question material to see the characteristics that can be identified as a type of question into the characteristics of compound intelligence as follows:

Table 1. Identifying characteristics of learning content

Compound intelligence (criteria)	Identification (criteria size)
Verbal	<ol style="list-style-type: none"> 1 Human objects (PRN) 2 The narrative of the problem is use of subject (NN) + predicate (VERB). 3 There is a foreign language term or uptake (FOREIGN) 4 There is an insert of monologue or dialogue content (DIALOGUE)
Logic	<ol style="list-style-type: none"> 1 Many involve objects in the form of numbers or variables (NUM). A threshold of 5 is set for objects in the form of numbers. 2 Use of a table as an object viewer (TABLE) 3 Use of logic variables (LOGIC) 4 There are mathematical symbols (MATH)
Visual	<ol style="list-style-type: none"> 1 There are visual content inserts (video/images) (IMG/VID) 2 Narrative tends to encourage visualizing shapes, objects, colors (SHAPE/COLOR) 3 Using images as question and answer (IMG) content 4 Make use of diagrams, charts (DIAGRAMS)
Natural	<ol style="list-style-type: none"> 1 Objects in the form of animals, plants, or natural objects (ANIMAL / PLANT / NATURE), Classification and grouping of plant and animal types (ANIMAL / PLANT) 2 The narrative of the question can be a natural phenomenon (NATURE)

The classification model was developed as a classifier to detect the type of compound intelligence in the question script. The corpus dataset used is a form of written test questions that will be labeled on each question. The data to be used as training data is taken from test questions on the online learning platform, which are then saved to the database in HTML notation format so that tags such as images.

Each instance in the question dataset is then manually labeled with reference to the characteristics of each class that it has defined on. The process of extracting information [16] with the aim of labeling the problem is as follows table 2:

Table 2. Illustration of identification and extraction of information

1	<p><p>RianPRN went VERB to the NN store of the NN writing tool VERB to buythe VERBPENIN PENPOIN. The price of 1 NUM fruit NN ballpoint pen NN Rp 1,750 NUM. IfLOG Rian PRN buys VERB 1 NUM dozen ballpoint pens and HePRN paysVERB 5 NUM billsNN five NUM thousands ofNUM. How much change does RianNN receive? </p></p>
---	--

From the example, it can be identified as follows:

- 1) Class verbal : Four mentions of the person's name object (PRN) were found, then there were complete sentence characteristics in the form of subject and predicate (PRN/NN + VERB), but no foreign language word (FOREIGN) was found.
- 2) Class visual : No image (IMG) or video (VID) attachment tags were found, and also no sentences were found describing the field visualization.
- 3) Logical classes : Found five words in the form of numbers or nominal mentions (NUM), found presuppositional sentences as characteristics of logical sentences (LOGIC).
- 4) Class natural :There is no mention of objects in the form of animals, plants, natural phenomena or other natural objects. Table 3 present of result of identification the problem label.

Table 3. Results of the identification of the problem label

Verbal		Visual		Logic		Natural	
PRN	4	IMG	0	NUM	5	ANIMAL	0
PRN/NN+VERB	1	VID	0	TABLE	0	PLANT	0
FOREIGN	0	SHAPE	0	LOG	1	NATURE TOPIC	0
class	1	class	0	class	1	class	0

From the results of the analysis of the problem, characteristics that indicate two types of compound intelligence have been identified, it is concluded that the question will be labeled verbally and logically. Table 4 present in the database, the question instance will be given a value of 1 in the class identified as positive and a value of 0 if vice versa.

Table 4. Multiclass dataset identification

Id	Question Content	Answer Content	Visual	Verbal	Logic	Natural
1	<p>Rian went to the store...</p>		0	1	1	0

C. Data Preparation

Figure 1 illustrates the stages of preparing the question manuscript dataset for processing into a model. At this stage, data preparation and model development are carried out in a python environment show in Figure 1. The stages of data preparation are as follows:

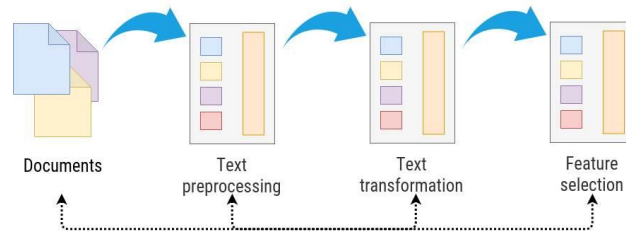


Figure 1. Dataset preparation illustration

- a Text preprocessing is a phase that consists of several processes. The first process is tokenize, which is to break words into a word matrix that is based on spaces, whitespace, tabs, newlines, and others. Furthermore, pay attention to the grammatical of the words used to correct imperfect or non-standard words to the standard word form. In general, this stage is carried out when handling the question dataset.
- b Text transformation also consists of several processes. The first process is to convert the case letters to non-capital. Then proceed with the process of stemming or lemmatization, which is to make word forms the basis in this case eliminating affixes and or suffixes. The stemming process also makes use of the Stemmer Factory function of the literary library.
- c Feature selections are a method of selecting significant features in a text set. There are two methods in feature selection, the first is to represent the word into a set of sequences and the frequency with which the word appears, and the second is to represent the word directly as a string. The first type is more widely used due to ease in implementation [22], [23]. The Term Frequency-Inverse Document Frequency (TF-IDF) method is commonly used in the implementation of word vector representation weighting and was adopted in this study.

D. Modeling

The number of datasets collected was 709 instances, with the format of the question text HTML notation. The purpose of storing data in HTML format is so that the form of tags that characterize the problem such as image tags, question attachments, and links listed in the question can be features that participate in the model training process. The classification method used is supervised learning [21], [24], [25], where each data instance has a label in the form of a type of compound intelligence, which is in the form of multi-label. Each instance of the question is then assigned to a label according to the type of compound intelligence that has been set. Because it is a multi-label type, each question can have more than one type of label. For example, a question instance can be a problem with a visual label and a logic label at the same time. This has an impact on the selection of machine learning methods in the development of models that detect more than one type of label in a question. The following figure shows the distribution of classes in the question dataset show in Figure 2.

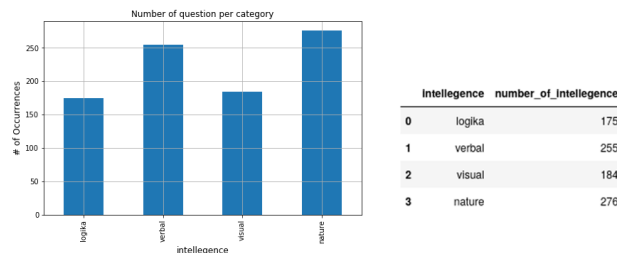


Figure 2. Distribution of labels on datasets

When analyzing the distribution of each class at fig. 2, it is known that the distribution of classes in the dataset is unbalanced (class-imbalanced) and a preliminary handling of the dataset will be carried out to balance the class by piloting several methods

E. Evaluation

At the evaluation & development stage, you will see the results of model development by conducting trials and comparing the achievements of each algorithm tried, by looking at the results of cross validation. According to surveys [23], [24] algorithms in text classification can be grouped into several types as in. From these various types of algorithms, this study will implement in the evaluation stage several types of algorithms in the probabilistic & naïve bayes, linear-based, and proximity-based groups.

F. Deployment

The next stage is to deploy the model into the built system. This process is the integration of the classifier model into the prototype of the developed application. Business objectives will be seen from the results of the achievement of implementing a simulation of an application prototype that should be able to work well on the model that has been developed.

IV. RESULT AND DISCUSSION

A. Strategi Class-Imbalanced

Figure The trial of class-imbalanced handling was carried out with several strategies, namely 1) classifier without patch method, 2) class weighting, 3) using resampling, 4) using ensemble, 5) class weighting and resampling, 6) class weighting and ensemble, 7) ensemble and resampling, 8) class weighting.

The combination of handling strategies is treated on the dataset with the technique of balancing the class biner values on each label. The following are the test results for each class with eight strategies that have been defined by applying a support vector machine (SVM) algorithm [26]–[31]with a linear kernel configuration show in table 5-8.

Table 5. Results of testing classification strategies on types of logical intelligence

Strategy	Precision	Recall	F1-score	Accuracy
[1] SVM	0,81	0,74	0,76	0,82
[2] SVM, CW	0,72	0,72	0,72	0,77
[3] SVM, RES	0,73	0,75	0,74	0,77
[4] SVM, ENS	0,84	0,75	0,78	0,84
[5] SVM, CW, ENS	0,82	0,75	0,77	0,83
[6] SVM, CW, RES	0,73	0,75	0,74	0,77
[7] SVM, RES, ENS	0,74	0,73	0,73	0,79
[8] SVM, CW, RES, ENS	0,78	0,74	0,75	0,81

Table 6. Results of testing classification strategies on types of visual intelligence

Strategy	Precision	Recall	F1-score	Accuracy
[1] SVM	0,91	0,89	0,90	0,92
[2] SVM, CW	0,86	0,90	0,88	0,90
[3] SVM, RES	0,88	0,91	0,89	0,91
[4] SVM, ENS	0,92	0,91	0,91	0,93

[5] SVM, CW, ENS	0,93	0,90	0,91	0,93
[6] SVM, CW, RES	0,88	0,91	0,89	0,91
[7] SVM, RES, ENS	0,90	0,91	0,91	0,92
[8] SVM, CW, RES, ENS	0,94	0,90	0,92	0,94

Table 7. Results of testing classification strategies on types of verbal intelligence

Strategy	Precision	Recall	F1-score	Accuracy
[1] SVM	0,84	0,82	0,83	0,86
[2] SVM, CW	0,79	0,80	0,79	0,82
[3] SVM, RES	0,79	0,80	0,79	0,82
[4] SVM, ENS	0,85	0,82	0,83	0,87
[5] SVM, CW, ENS	0,81	0,81	0,81	0,85
[6] SVM, CW, RES	0,79	0,80	0,79	0,82
[7] SVM, RES, ENS	0,79	0,80	0,79	0,82
[8] SVM, CW, RES, ENS	0,83	0,82	0,83	0,86

Table 8. Results of testing classification strategies on types of natural intelligence.

Strategy	Precision	Recall	F1-score	Accuracy
[1] SVM	0,94	0,91	0,92	0,93
[2] SVM, CW	0,91	0,91	0,91	0,92
[3] SVM, RES	0,92	0,90	0,91	0,92
[4] SVM, ENS	0,92	0,91	0,92	0,92
[5] SVM, CW, ENS	0,91	0,91	0,91	0,92
[6] SVM, CW, RES	0,92	0,90	0,91	0,92
[7] SVM, RES, ENS	0,90	0,90	0,90	0,91
[8] SVM, CW, RES, ENS	0,94	0,92	0,93	0,94

The test results of Table 5-8 provide values by establishing the following strategy show in Table 9.

Table 9. Imbalanced-class strategy test summary

Class	Strategy		
	Class-Weighting	Resampling	Ensamble
Logic	X	X	√
Verbal	X	X	√
Visual	√	√	√
Natural	√	√	√

B. Classifier

Based on the next class-imbalances handling strategy [27], training data is carried out to divide the dataset into two parts with random instance selection, namely 0.8 as training data and the remaining 0.2 as test data to see the accuracy of the built model. Complete test results as shown in the Table 12 by implementing support vector machines (SVM),

naïve bayes (NB), logistic regression (LR), neural networks (NN), K-nearest neighbors (KNN), and stochastic gradient descent (SGD). Because the classification model is multi-label, model development can be done by applying different algorithmic methods to each classifier model. A summary of the best results is as in the following table 10.

Table 10. Summary of test results with the highest scores

Types of intelligence	Method algorithm	Score		
		F1-score Macro Avg	F1-score Weighted Avg	Accuracy
Logic	Support vector machine (SVM)	0,78	0,83	0,84
Verbal	Logistic regression (LR)	0,84	0,87	0,87
Visual	Logistic regression (LR)	0,91	0,93	0,93
Natural	Neural network (NN)	0,93	0,93	0,93

V. CONCLUSION

The implementation of text mining on text classification questions provides multi-class modeling in which a different classifier can be implemented in each class. The use of datasets that utilize questions on online learning platforms results in unbalanced data classes, so hyperparameter experiments are implemented to get the best handling strategy. Implementation and evaluation of the classification model will be carried out to continue the implementation in the class.

In this research, we succeeded in creating a proposed classification model for exam question papers based on four types of multiple intelligences with a combination of the best methods produced by the SVM method, namely for the logical type accuracy 0.85, logistic regression visual type accuracy 0.87, logistic regression verbal type accuracy 0.93, and neural network natural type accuracy 0.96.

With excellent results from the personalized written test experiment, elementary school level students experienced an increase in their average score of 3.06 points compared to before the personalized test was carried out. This will greatly influence students to improve their learning which is of course adapted to their natural talents. In the future, it is hoped that it can also be used to create analytical models based on musical, kinesthetic, interpersonal and intrapersonal intelligence with the aim of expanding the range of personalization. So that if it can cover all types of intelligence, a model will be found to make personalized tests for students more appropriate during learning or testing, of course this will make it easier to create a road map for appropriate and effective curriculum design for children's education in Indonesia. This model uses the research object of elementary school students, where if intelligence tendencies can be determined, it will make it easier for teachers, parents and the government to take the right steps to implement the right educational model for each student.

REFERENCES

- [1] Z. Lu, "Instruction and Evaluation of University Physics Experiment under the Theory of Multiple Intelligences," *Proc. - 2015 Int. Symp. Educ. Technol. ISET 2015*, pp. 79–83, Mar. 2016, doi: 10.1109/ISET.2015.24.
- [2] E. Kusniati, "STRATEGI PEMBELAJARAN BERBASIS MULTIPLE INTELLIGENCES".
- [3] M. Pocinho and C. Mendes, "Primary School Children's Multiples Intelligences Assessment," *Psicol. Teor. e Pesqui.*, vol. 37, pp. 1–9, 2021, doi:

- 10.1590/0102.3772E37304.
- [4] X. Liu, J. Gu, and L. Zhao, "Promoting primary school students' creativity via reverse engineering pedagogy in robotics education," *Think. Ski. Creat.*, vol. 49, p. 101339, Sep. 2023, doi: 10.1016/J.TSC.2023.101339.
 - [5] V. Prain et al., "Personalising Learning," *Adapt. to Teach. Learn. Open-Plan Sch.*, pp. 43–58, 2014, doi: 10.1007/978-94-6209-824-4_3.
 - [6] B. Said, L. Cheniti-Belcadhi, and G. El Khayat, "An ontology for personalization in serious games for assessment," *Proc. - IEEE 2nd Int. Conf. Artif. Intell. Knowl. Eng. AIKE 2019*, pp. 148–154, Jun. 2019, doi: 10.1109/AIKE.2019.00035.
 - [7] M. Ren, Z. Wang, and X. Yu, "Personalized federated learning: A Clustered Distributed Co-Meta-Learning approach," *Inf. Sci. (Ny)*, vol. 647, p. 119499, Nov. 2023, doi: 10.1016/J.INS.2023.119499.
 - [8] B. Bray and K. McClaskey, "Mid-Pacific Institute 1:1 iPad Program Personalization vs Differentiation vs Individualization," pp. 1–13, 2014, [Online]. Available: <http://www.ed.gov/technology/draft-netp-2010/individualized-personalized-differentiated-instruction>
 - [9] J. S. Akbar et al., *Model dan Metode Pembelajaran Inovatif (Teori dan Panduan Praktis)*, vol. 1. 2023.
 - [10] K. B. Zafman, M. L. Riegel, L. D. Levine, and R. F. Hamm, "An interactive childbirth education platform to improve pregnancy-related anxiety: a randomized trial," *Am. J. Obstet. Gynecol.*, vol. 229, no. 1, pp. 67.e1-67.e9, Jul. 2023, doi: 10.1016/J.AJOG.2023.04.007.
 - [11] S. M. Larson, T. Cox, T. Carelock, and E. S. DeMeyer, "Interactive Education: One Way to Measure Success," *Transplant. Cell. Ther.*, vol. 30, no. 2, pp. S73–S74, Feb. 2024, doi: 10.1016/J.JTCT.2023.12.119.
 - [12] D. Li, "An interactive teaching evaluation system for preschool education in universities based on machine learning algorithm," *Comput. Human Behav.*, vol. 157, p. 108211, Aug. 2024, doi: 10.1016/J.CHB.2024.108211.
 - [13] C. Saul and H. D. Wuttke, "E-assessment meets personalization," *IEEE Glob. Eng. Educ. Conf. EDUCON*, pp. 200–206, 2013, doi: 10.1109/EDUCON.2013.6530106.
 - [14] K. A. McCusker, J. Harkin, S. Wilson, and M. Callaghan, "Intelligent assessment and content personalisation in adaptive educational systems," *2013 12th Int. Conf. Inf. Technol. Based High. Educ. Training, ITHET 2013*, 2013, doi: 10.1109/ITHET.2013.6671025.
 - [15] D. E. Benchoff, M. P. Gonzalez, and C. R. Huapaya, "Personalization of Tests for Formative Self-Assessment," *Rev. Iberoam. Tecnol. del Aprendiz.*, vol. 13, no. 2, pp. 70–74, May 2018, doi: 10.1109/RITA.2018.2831759.
 - [16] L. T. M. Blessing and A. Chakrabarti, "DRM, a design research methodology," *DRM, a Des. Res. Methodol.*, pp. 1–397, 2009, doi: 10.1007/978-1-84882-587-1/COVER.
 - [17] "DRM, a Design Research Methodology | SpringerLink." <https://link.springer.com/book/10.1007/978-1-84882-587-1> (accessed Mar. 16, 2024).
 - [18] G. Zhao, L. Yang, J. Li, J. Chu, and Y. Qi, "Design and Implementation of a Teaching Verbal Behavior Analysis Aid in Instructional Videos," *Proc. - 2022 4th Int. Work. Artif. Intell. Educ. WAIE 2022*, pp. 1–5, 2022, doi: 10.1109/WAIE57417.2022.00008.
 - [19] H. Karso, "Pernyataan dan Kata Hubung Pernyataan Majemuk".
 - [20] C. G. Interiano, D. P. T. Tkacik, J. L. Dahlberg, and D. J. H. Lim, "Authentic Knowledge, Learning Outcomes, and Professional Identity: A Mixed-Methods Study of a Successful Engineering Course," *Proc. - Front. Educ. Conf. FIE*, vol. 2018-October, Jul. 2018, doi: 10.1109/FIE.2018.8659152.

- [21] C. Liu and Z. Chao, "Supervised learning and unsupervised learning on music data with different genres," *Proc. - 2021 IEEE 7th Int. Conf. Big Data Intell. Comput. DataCom 2021*, pp. 7–12, 2021, doi: 10.1109/DATACOM53700.2021.00008.
- [22] A. R. Mishra, V. K. Panchal, and P. Kumar, "Extractive Text Summarization-An effective approach to extract information from Text," *Proc. 4th Int. Conf. Contemp. Comput. Informatics, IC3I 2019*, pp. 252–255, Dec. 2019, doi: 10.1109/IC3I46837.2019.9055636.
- [23] C. C. Aggarwal and C. X. Zhai, "A survey of text classification algorithms," *Min. Text Data*, vol. 9781461432234, pp. 163–222, Aug. 2012, doi: 10.1007/978-1-4614-3223-4_6/COVER.
- [24] F. Gurcan, "Multi-Class Classification of Turkish Texts with Machine Learning Algorithms," *ISMSIT 2018 - 2nd Int. Symp. Multidiscip. Stud. Innov. Technol. Proc.*, Dec. 2018, doi: 10.1109/ISMSIT.2018.8567307.
- [25] R. Saravanan and P. Sujatha, "A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification," *Proc. 2nd Int. Conf. Intell. Comput. Control Syst. ICICCS 2018*, pp. 945–949, Jul. 2018, doi: 10.1109/ICCONS.2018.8663155.
- [26] L. Li, S. Ma, and Y. Zhang, "Optimization algorithm based on genetic support vector machine model," *Proc. - 2014 7th Int. Symp. Comput. Intell. Des. Isc. 2014*, vol. 1, pp. 307–310, Mar. 2015, doi: 10.1109/ISCID.2014.99.
- [27] X. Zhang and B. G. Hu, "A new strategy of cost-free learning in the class imbalance problem," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 12, pp. 2872–2885, Dec. 2014, doi: 10.1109/TKDE.2014.2312336.
- [28] Z. Lai, G. Liang, J. Zhou, H. Kong, and Y. Lu, "A joint learning framework for optimal feature extraction and multi-class SVM," *Inf. Sci. (Ny.)*, vol. 671, p. 120656, Jun. 2024, doi: 10.1016/J.INS.2024.120656.
- [29] Z. Tong *et al.*, "Coal structure identification based on geophysical logging data: Insights from Wavelet Transform (WT) and Particle Swarm Optimization Support Vector Machine (PSO-SVM) algorithms," *Int. J. Coal Geol.*, vol. 282, p. 104435, Feb. 2024, doi: 10.1016/J.COAL.2023.104435.
- [30] C. Ding, Y. Xia, Z. Yuan, H. Yang, J. Fu, and Z. Chen, "Performance prediction for a fuel cell air compressor based on the combination of backpropagation neural network optimized by genetic algorithm (GA-BP) and support vector machine (SVM) algorithms," *Therm. Sci. Eng. Prog.*, vol. 44, p. 102070, Sep. 2023, doi: 10.1016/J.TSEP.2023.102070.
- [31] Y. Yu *et al.*, "Quantitative analysis of multiple components based on support vector machine (SVM)," *Optik (Stuttg.)*, vol. 237, p. 166759, Jul. 2021, doi: 10.1016/J.IJLEO.2021.166759.