JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY
https://e-journal.uum.edu.my/index.php/jict

How to cite this article:
Muhaimin, A., Wibowo, W., & Riyantoko, P. A. (2023). Multi-label classification using vector generalized additive model via cross-validation. *Journal of Information and Communication Technology, 22*(4), 657-673. https://doi.org/10.32890/jict2023.22.4.5

# Multi-label Classification Using Vector Generalized Additive Model via Cross-Validation

[1]Amri Muhaimin, [*2]Wahyu Wibowo &
[3]Prismahardi Aji Riyantoko
[1]Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia
[2]Department of Business Statistics, Institut Teknologi Sepuluh Nopember, Indonesia
[3]Department of Information and Communication Systems, Okayama University, Japan

[1]amri.muhaimin.stat@upnjatim.ac.id
[*2]wahyu_w@statistika.its.ac.id
[3]pnai2m3s@s.okayama-u.ac.jp
*Corresponding author

Received: 14/7/2023 Revised: 16/10/2023 Accepted: 17/10/2023 Published: 25/10/2023

## ABSTRACT

Multi-label classification is a unique challenge in machine learning designed for two targets with each containing one or multiple classes. This problem can be resolved using several methods, including the classification of the targets individually or simultaneously. However, most models cannot classify the target simultaneously, and this is not expected to happen in the modeling rule. This study was conducted to propose a novel solution in the form of a Vector Generalized Additive Model Using Cross-Validation (VGAMCV) to address these problems. The proposed method leverages the Vector

Generalized Additive Model (VGAM), which is a semi-parametric model combining both parametric and non-parametric components as the underlying base model. Cross-validation was also applied to tune the parameters to optimize the performance of the method. Moreover, the methodology of VGAMCV was compared with a tree-based model, Random Forest, commonly used in multi-label classification to evaluate its effectiveness based on fourteen metric scores. The results showed positive outcomes as indicated by 0.703 average accuracy and 0.601 Area Under Curve (AUC) recorded, but these improvements were not statistically significant. Meanwhile, the method offered a viable alternative for multi-label classification tasks, and its introduction served as a contribution to the expanding repertoire of methods available for this purpose.

**Keywords:** Classification, cross-validation, generalized model, multi-label, semi-parametric.

## INTRODUCTION

Classification is a prevalent task in the field of data analysis and is traditionally closely associated with logistic regression in statistical modeling. Several classification methods are usually applied in different domains, including science, business, and social science, due to their effectiveness in solving many problems (Chowdhury & Schoen, 2020). For example, some models have been used in the healthcare sector to diagnose patients with specific conditions, such as diabetes and cardiovascular diseases, in order to offer valuable insights and ensure effective treatment decisions (Alic et al., 2017). However, these models often encounter challenges such as overfitting, underfitting, and the perfect case scenario.

Overfitting occurs when a model performs effectively on the training data but fails to generalize to new and unseen data, while underfitting happens when it cannot capture the complexities of the underlying data. The perfect case refers to a situation where a model achieves 100% accuracy, leading to concerns about potential data leakage or unrealistic performance (Muhamedyev et al., 2015). Several regularization methods, feature selection, and cross-validation strategies have been used to address these challenges in order to

optimize classification models and ensure their generalizability by balancing complexity and performance, as well as avoiding underfitting and overfitting scenarios.

The remaining aspects of this study contain a discussion of the method theoretically and a brief description of the multi-label classification problem in the next section. This is followed by an explanation of the methodology and the solution designed. The proposed method is later compared with the method developed by Charte and Charte (2015) in the R package called Multilabel Dastasets in R (MLDR) using random forest as the base model and one-by-one rule to classify the target instead of simultaneously. The following section presents results and discussion, while the last focuses on the conclusion and future works.

## RELATED WORKS

Classification is categorized into several types, including binary, multiclass, and multi-label. The binary focuses on only one target binary class, such as zero or one and male or female. Multiclass is similar to binary, but the class in the target is non-binary, such as the famous iris dataset. Furthermore, classification also correlated with multivariate analysis, where the response variable or target is usually more than one column, thereby indicating the need to predict two targets simultaneously. This multivariate assumes a correlation between target features such as $y_1$ correlates to $y_2$ (Read et al., 2014). However, this assumption is limited to normal distribution data and correlated target features. It is difficult to apply this multivariate analysis when the feature is not correlated, and the data is not normally distributed. This method traditionally relies on hypothesis testing, and this has the ability to limit its flexibility when dealing with data that does not follow a normal distribution. Multivariate poses a challenge when analyzing datasets with non-normal distributions, and this means there is a need for alternative methods to handle such scenarios while still producing reliable results. This limitation needs to be handled using a method that is able to effectively analyze non-normally distributed data, provide robust outcomes, offer flexibility in accommodating different data distributions, and yield accurate and meaningful insights from multivariate analysis.

The use of methods that do not rely heavily on strict assumptions about data distribution allows the exploration of alternatives, offering greater flexibility and adaptability to diverse datasets. These include the non-parametric or semi-parametric models that capture complex relationships and patterns in the data, even in non-normal distributions. Moreover, the application of more robust methods against deviations from normality allows a wider range of data types and distributional characteristics to be analyzed effectively. The methods open new possibilities to gain insights and make informed decisions based on multivariate analysis, even in scenarios where traditional methods have the potential to fall short. It is important to have a method that can handle non-normally distributed data in multivariate analysis is crucial. Therefore, the adoption of more flexible strategies and resilience to deviations from normality makes it possible to overcome limitations and obtain meaningful results across a broader range of datasets and real-world scenarios.

Some methods have been widely used for multi-label classification, such as multivariate terms, which predict all targets simultaneously based on order, with the focus on the first target followed by the second or vice versa. For example, Read et al. (2014) applied a meta-label to classify all targets simultaneously, and the method tackled nonlinearities effortlessly. Pan et al. (2020) also proposed Feature-level Attention in the clinical text using Natural Language Processing (NLP)-based data. The method used a distance calculation to classify the data and produced more training time than conventional methods. It is also possible to use probabilistic methods to solve multi-label classification problems, specifically for protein function, music, semantic scene, and several others (Kommu et al., 2014). The process is usually based on the application of logistic regression and the nearest neighbor classifier as the base model to classify the problem. However, all the methods previously mentioned do not have the ability to capture the correlation between targets. A class alignment method was proposed by Taha et al. (2021) through the development and comparison of the Multilabel Over-Sampling And Under-Sampling With Class Alignment (ML-OUSCA) model with K-Means and K-Nearest Neighbor to over or under-sample multi-label, but it was only able to classify targets individually and considered a waste in a situation consisting of several label. Therefore, this study proposed a Vector Generalized Additive Model (VGAM) Using Cross-Validation (VGAM-UCV) to solve this problem.

VGAM-UCV is a combination of cross-validation and VGAM to avoid underfitting and overfitting in the process of conducting multi-label classification. Cross-validation is a method normally used in machine learning and statistical modeling to assess the performance and generalization capability of a predictive model. The process is usually through the splitting of a dataset into multiple subsets, training and evaluating the model on different combinations of these subsets. Meanwhile, VGAM is a semi-parametric model that can be used for linear and non-linear data but is considered vulnerable to overfitting or underfitting (Yee et al., 1996). The workflow in this study was initiated from data pre-processing to the creation of training-split and cross-validation, optimization of the parameters, selection of the best model, and evaluation. The aim was to determine the performance of VGAM-UCV by comparing it with the tree-based model in a multi-label case. Moreover, Mohammed et al. (2020) were observed to have applied Deep Learning to solve NLP multi-label cases, but the model was considered very complex. This led to the usage of data from the UCI dataset and cross-sectional classification cases in this study.

## MATERIALS AND METHODS

This section provides an overview of the materials and methods used in this study with a focus on several key components, including explanations of multi-label classification, VGAM, metric evaluation, and methodology. The primary objective of this investigation is to achieve high accuracy and identify the best model for multi-label classification prediction. Moreover, even though the assessment of the goodness of fit for the model is typically crucial, other factors are prioritized due to the unique nature of multi-label classification. VGAM framework used follows a semi-parametric method, and this eliminates the need for hypothesis testing on the parameters of the model. This methodology is leveraged in this study to enhance the predictive capabilities of a multi-label classification system.

### Multi-label Classification

Multi-label explanations make it possible to address the scenarios where multiple labels are assigned to a single instance. This condition is particularly valuable in complex classification tasks where objects or samples possess multiple attributes or belong to

multiple categories at the same time. In a situation where data have responses notated with $\{x_i \mid i \in 1,\ldots,n\}$ and target features y, this can lead to the categorization of the data into two classes $\{y_i \mid j \in 1,2\}$. The intention is to predict the class of the new data as 1 or 2, and several classification methods, specifically the single and multi-label, have been designed to achieve this over time. The single aspect works by categorizing the data into some classes, but the target feature is formed as a one-dimensional column vector (Venkatesan & Er, 2014). Meanwhile, the multi-label aspect has more than one column and is usually referred to as multivariate or bivariate analysis in statistics because the response variable commonly notated by $y$ is more than 1. It is important to note that multi-label is different from the single-label due to the existence of more than one target variable, with each containing a single or multiclass class. This makes it easy to apply traditional or conventional machine-learning methods in solving this problem (Wang et al., 2017).

Multi-label classification is a versatile method normally applied in different domains such as text, protein function, and music categorization. Several foundational methods in text categorization use feature selection in multi-label classification to enhance accuracy (Rahmawati & Khodra, 2015). Meanwhile, the challenges associated with high-dimensional data extend beyond the complexities inherent to multi-label classification. The nature of the target variables is also important because the problem of those considered categorical is usually linked to multi-label classification, while numerical ones are either bivariate or multivariate regression. The objective of multi-label classification is to assign multiple labels or categories to a given input in order to ensure a more nuanced understanding and representation of the underlying data. It has become a valuable tool to facilitate the extraction of meaningful insights from complex and diverse data sources due to the increasing availability of large datasets and advancements in machine learning algorithms.

## VECTOR GENERALIZED ADDITIVE MODEL

The correlation between the response and predictor variables is sometimes not linear or non-linear. The linear model creates a one-dimensional relationship between the predictor and response variable based on the equation $y=x\beta+\varepsilon$ and this is known as the

linear regression. The equation can be used for linear relationships but is not applicable to those considered non-linear. This led to the development of a generalized additive model to accommodate the non-linear relationship between the response and predictor variable. This is a semiparametric method that uses a smoother to handle the nonlinearity relationship in the data and is represented as indicated in the following Equation 1.

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots + f_d(x_d), \quad (1)$$

where, $y$ is the response variable while the predictor has a smoothing function, $f_i$, $1 \leq i \leq d$, in each variable, indicating the $y$ variable is not linearly related to $x$. The smoothing function is usually penalized splines, specifically in R software. The replacement of the $d$ in Equation 1 with 2 led to the formulation of the model as presented in the following Equation 2.

$$y = \beta_0 + \beta_1 x_1 + \sum u_{1k} z_{1k}(x_1), \quad (2)$$

where, $z_{1k}(\cdot)$ is an O'Sullivan spline basis over x variable, and the number of penalization is also applicable to the $u_{ik}$. Equation 2 shows a change in the response variable $y$ based on the $x$ variable. Moreover, the analysis becomes VGAM when the response variable is multivariate, and this changes Equation 2 to the following.

$$\begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix} = \begin{bmatrix} \beta_{10} + \beta_{11} x_{1i} + f_{12}(x_{2i}) \\ \beta_{20} + \beta_{21} x_{2i} + f_{22}(x_{2i}) \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix}. \quad (3)$$

Equation 3 shows that the model contains two targets and two variables with one smoothing function, $f(\cdot)$, while the ε stands for the error in each target model (Harezlak et al., 2018).

A generalized model offers several advantages, and these include the adjustment of the smoothing function and the use of different distribution methods based on the characteristics and nature of the data. For example, the data consisting of discrete values such as gender can be analyzed using the binomial distribution or other alternatives such as Poisson or negative binomial distributions. This flexibility allows the effective application of the model to diverse data types.

The benefits of the generalized model extend beyond the target variable, as indicated by the possibility of applying the smoothing function to the features of the data, which improves its ability to have additional flexibility and handle different data distributions. This means the inclusion of an appropriate smoothing function and selection of the most suitable distribution for the given data can allow the model to effectively capture the underlying patterns and dependencies within the dataset. This adaptability and versatility increase the importance of the model in different applications and also enable robust analysis and modeling in diverse domains.

## The Proposed Method

This section was used to present the data and processes applied to conduct this study. The first step was to pre-process the data before the analysis and modeling, and this was achieved by tuning and cross-validating the parameters to select the best model for the test. Cross-validation was defined as the new way to address and tune the parameters of VGAM and this was achieved using the data on abalone or marine snails. The purpose was to classify the data on the gender and 'rings' which were found to be both binary classifications because the 'gender' consisted of male and female while the 'rings' were zero and one. All the features used such as length, diameter, height, whole height, shucked weight, viscera weight, and shell weight were numeric. Moreover, the workflow for the model presented in Figure 1 shows that the data were extracted and first pre-processed by checking the missing and duplicate values for appropriate corrections. This was followed by the training of the model using VGAM via cross-validation (VGAM-CV) to tune the parameters. The next step was to check or evaluate the model, and the final was to test and compare the data with the benchmark.

**Figure 1**

*Modeling Scenario for VGAM with Cross-Validation*

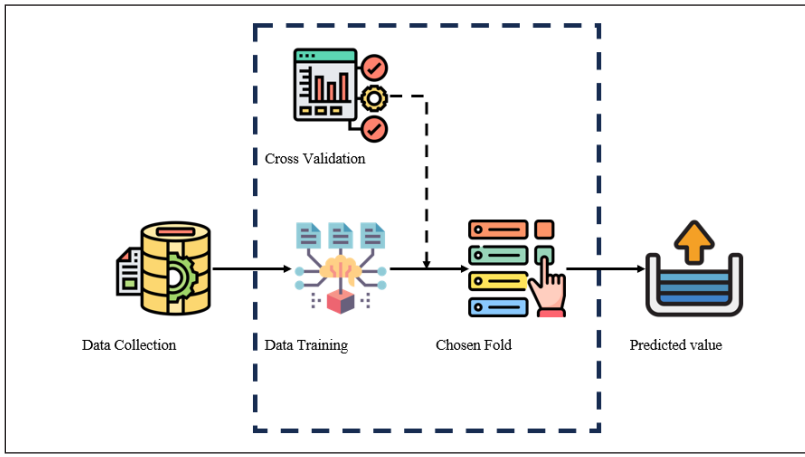



Figure 1 shows that the data containing two targets, with each having a binary class, were stored in databases, extracted, and split into training and test sets for the modeling process. The data were checked for missing and duplicate values for appropriate transformation before the initiation of the training process. Moreover, cross-validation was applied during the training process to assist the model experiencing underfitting or overfitting as well as to determine the best training set to classify the data. The selection of the best fold with the highest accuracy in classifying the data was followed by the development and evaluation of the predicted value using the test set. This method was applied to produce a robust result because the generalized additive model was designed to handle multivariate or multi-response analysis. The final step was used to test and compare the data with the tree-based Random Forest used as the benchmark.

The Random Forest is widely known as the ensemble method and has been successfully applied in different areas to classify multi-label cases. For example, it was used by Tabia (2019) to solve multi-label classes and produced a positive result. Sharma and Mehrotra (2018) also compared several methods of solving multi-label cases, and Random Forest was found to have produced 96% accuracy. This led to its application as the benchmark in this study based on 14 14-metric score evaluations due to the inability of one metric to measure the performance of VGAM-CV relative to the benchmark.

**Metric Evaluation**

The metrics usually used to evaluate classification are accuracy, sensitivity, specificity, recall, precision, and F1 score, but the method for multi-label cases is different. The aim of this study was to evaluate the whole responses simultaneously, and this showed the possibility of having more than one of these metrics with the implementation of multi-label classification. Novita Sari et al. (2021) measured performance using the accuracy from cross-validation and the geometric mean of each metric, including the F1, recall, precision, and accuracy, but the results were not very representative. This showed the need for special calculations in the process of measuring the performance of the model applied to multi-label cases.

The multi-label classification was observed to usually have more than one confusion matrix, and this indicated the existence of three different confusion matrices for the three labels identified in this study. Therefore, there was the need to determine the aggregate, such as the macro, micro, weighted, and sample average, using the formulas presented in the following Table 1.

**Table 1**

*The Formula of Metric Score for Multi-label Classification*

| Metric | Formula |
|---|---|
| Macro Average | $\frac{\Sigma_{y_i, i\epsilon 1,..,n} Score}{n}$ |
| Micro Average | $\frac{\Sigma_{y_i, i\epsilon 1,...,n} TP}{\Sigma_{y_i, i\epsilon 1,...,n} TP + \Sigma_{y_i, i\epsilon 1,...,n} FP}$ |
| Weighted Average | $\frac{\Sigma_{y_i, i\epsilon 1,...,n} (Score)(Support)}{\Sigma_{y_i, i\epsilon 1,...,n} Support}$ |
| Samples Average | Use sampling to calculate the score. |

The macro average is often considered the optimal method to calculate metric evaluations because it assigns equal weight to each score, thereby allowing the computation of the average for several metrics such as accuracy, precision, recall, and F1-score to determine the comprehensive overview of the model performance. Meanwhile, the micro average calculates the average by considering each sample

individually, making it suitable for all metrics except accuracy. It also considers true and false positive values to generate the desired outcome. The selection of an appropriate averaging method when evaluating classification models depends on the specific requirements and objectives of the task. For example, the macro provides a balanced assessment by treating all scores equally, while the micro accounts for individual samples, with accuracy being the exception due to its distinct calculation requirements. This shows that the selection of the right evaluation strategy ensures a comprehensive understanding of the effectiveness of the model and aids in making informed decisions for further improvements or applications.

## RESULTS AND DISCUSSION

The data were trained on some fold options selected using the grid search. An attempt was made to combine from three to five-folds, and the result from the splitting process showed that the five folds had robust test and training set accuracy. Meanwhile, some parameters defaulted and were tagged with the smoothing function but not tuned because the focus was on determining the best model in prediction and not the significant parameter. The pattern of the data was assessed, and a smoothing function was added to those having a non-linear correlation between the variables and the target. Moreover, several combinations were tried for the data, but 80% training and 20% testing were found to be the best choice for this case. Normal standard transformation was applied to the feature in order to ensure more comparability because the variables had different measurements. The setting of all the parameters was subsequently followed by the running and evaluation of the model as well as its comparison with the benchmark.

Random Forest, used as the benchmark for multi-label cases, was developed in R software through a package called *MLDR*. Random Forest is one of the best machine learning algorithms, specifically in tabular data, due to the ability of its ensemble concept to ensure extreme robustness. However, the algorithm classifies the data as one target or one response, and this means the targets are usually categorized one by one. This led to the assumption of its usage as a benchmark for the proposed method designed to classify the target simultaneously.

The parameters used were tuned to determine the best smoothing value for the model, while cross-validation was applied to determine the

existence of overfitting, and the results are presented in the following Table 2. Moreover, only the accuracy was measured because it represented the other metrics. It was discovered that the VGAM-CV method was not overfitting or underfitting because there was no significant difference between the values recorded for the training and testing in each fold. This led to the conclusion that the model was robust or stable enough to classify the problem with 70% accuracy.

**Table 2**

*Five Folds of Cross-Validation Result*

| Metric | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Fold 5 |
|---|---|---|---|---|---|
| Accuracy Train | 0.720 | 0.720 | 0.710 | 0.720 | 0.720 |
| Accuracy Test | 0.700 | 0.710 | 0.730 | 0.700 | 0.710 |

VGAM was observed to be very robust, as expected, and this showed that the generalized model worked effectively through the data and had the ability to fit the target distribution, including binomial, Poisson, or normal. The raw or probability value was obtained from the model, and the output was observed to have originated from the sigmoid function in the range of zero to one because binomial distribution was applied. Moreover, there was no significant reduction in each fold for both the training and test sets. It was also discovered that the model was able to overcome the data not balanced in proportion to the target. A splitting rule of 0.5 was applied as the probability value, and the model was developed after cross-validation using all the data obtained for the parameters. The model was later trained and tested directly using the methods specified, and the results are presented in the following Table 3.

**Table 3**

*Comparison of VGAM-CV to the Random Forest*

| Method | VGAM-CV | MLDR: Random Forest |
|---|---|---|
| **Average Accuracy** | **0.703** | 0.702 |
| **Average Precision** | **0.967** | 0.960 |
| **Macro AUC** | **0.601** | 0.590 |
| Macro F-measure | 0.772 | **0.792** |

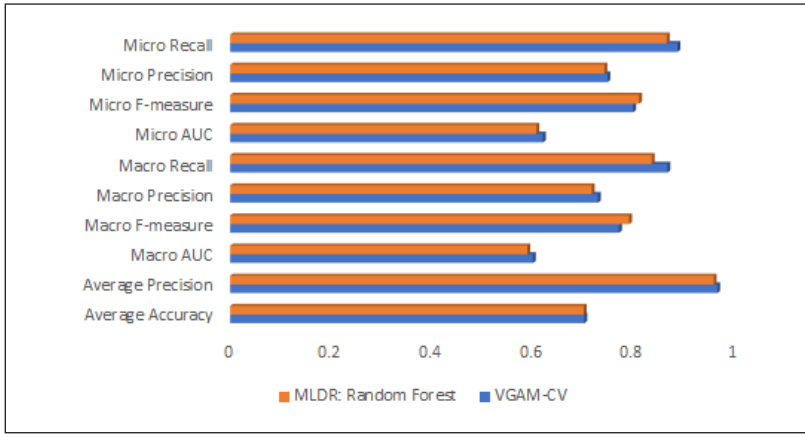
(continued)

| Method | VGAM-CV | MLDR: Random Forest |
|---|---|---|
| **Macro Precision** | **0.730** | 0.718 |
| **Macro Recall** | **0.868** | 0.837 |
| **Micro AUC** | **0.621** | 0.608 |
| Micro F-measure | 0.800 | **0.812** |
| **Micro Precision** | **0.749** | 0.743 |
| **Micro Recall** | **0.888** | 0.867 |

Table 3 and Figure 2 provide a comprehensive overview of the performance metrics compared between VGAM-CV and MLDR: Random Forest. The results showed that VGAM-CV performed better in most of the metrics except F-measure, and the difference was also observed to be relatively minor. The accuracy metric was used to determine the ratio of correct predictions to the total data points, and this provided insight into the general correctness of the model.

**Figure 2**

*Modeling Scenario of VGAM with Cross-Validation*




The area under curve (AUC) metric is the discriminative power of the model, and higher values represent superior performance. Another significant metric is the Average Precision, which is the mean of precision values recorded across the target features and focuses on the incisiveness of the model in predicting the classes. Moreover, the micro and macro metrics share similar meanings with common scores but have distinct formulations, as explicitly stated in Table 1.

The Random Forest Ensemble method is widely recognized for its effectiveness in handling structured data. Its basic concept of combining multiple models to process and classify data has shown impressive performance across different domains. Meanwhile, the proposed method focuses on using a single model for classification, making it difficult to compare the two methodologies directly. This means it lacks the benchmark for comparison, but the results are noteworthy and show a significant improvement compared to previous methods. The proposed method also offers cost advantages due to its relative simplicity and outperforms the MLDR in certain metrics despite the lower complexity, thereby indicating its effectiveness in classifying the data.

Some limitations were identified in the process of conducting this study, and the first was the fact that the dataset used only had numerical features due to inherent restrictions. This restricted the opportunity to explore other types of features that could potentially enhance classification performance. Second, there was no suitable benchmark for multi-label classification, and this led to the use of MLDR: Random Forest from the R package, which limited the scope for comparison and could cause bias. Future studies should address these limitations to refine further and validate the proposed methodology.

## CONCLUSION

In conclusion, several experiments were conducted to measure the performance of the proposed VGAM-CV for multi-label classification compared to the tree-based model, and the result showed the proposed method was considerable enough to handle multi-label data. This was important considering the existence of several multi-label data such as topics in the news, patients with some diagnostic, and others. Classification using the model was found to be relatively easy because the individual target was classified, indicating its effectiveness and efficiency. However, only one model was used to classify multiple responses in the data, thereby indicating the need for another to categorize the targets one by one, and this was considered ineffective and inefficient due to the time and memory to be consumed. VGAM-CV was identified as a semi-parametric method commonly used in

multivariate cases and considered a model-based method. However, the model was not necessary for this study because the objective was to accurately classify multi-label. An error was recorded in the process, but it was not required, like regression or time series analysis methods such as Autoregressive Moving Average (ARIMA). This VGAM-CV was found to be beneficial due to its ability to handle nonlinearity.

The smoothing function was able to adapt effectively to tackle nonlinearity terms, and this was the reason VGAM-CV was believed to be considerable compared to the tree-based model. Moreover, Random Forest used an average of major voting due to the generation of several more models than the ensemble, while VGAM-CV used a single model for the prediction and considered an alternative for multi-label classification. The method also used quite simple data, and the features were not much or presented in high dimensions. It was recommended that future studies test VGAM-CV against high-dimensional data, such as multi-label classification in NLP and measure the performance of this method on text data, specifically in the classification of topic news.

## ACKNOWLEDGMENT

This study was funded through the litdimas grant from Universitas Pembangunan Nasional "Veteran" Jawa Timur. The authors appreciate the higher education minister for their support and dedication to the study.

## REFERENCES

Alic, B., Gurbeta, L., & Badnjevic, A. (2017). Machine learning techniques for classification of diabetes and cardiovascular diseases. *2017 6th Mediterranean Conference on Embedded Computing (MECO),* 1–4. https://doi.org/10.1109/MECO.2017.7977152

Charte, F., & Charte, D. (2015). Working with multi-label datasets in R: The MLDR Package. *The R Journal, 7*(2), 149. https://doi.org/10.32614/RJ-2015-027

Chowdhury, S., & Schoen, M. P. (2020, October 2). Research Paper Classification using Supervised Machine Learning Techniques. *2020 Intermountain Engineering, Technology and Computing, IETC 2020.* https://doi.org/10.1109/IETC47856.2020.9249211

Harezlak, J., Ruppert, D., & Wand, M. P. (2018). Semiparametric Regression with R. Springer New York. https://doi.org/10.1007/978-1-4939-8853-2

Kommu, G. R., Trupthi, M., & Pabboju, S. (2014). A novel approach for multi-label classification using probabilistic classifiers. *2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)*, 1–8. https://doi.org/10.1109/ICAETR.2014.7012929

Mohammed, H. H., Dogdu, E., Gorur, A. K., & Choupani, R. (2020). Multi-label classification of text documents using deep learning. *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, 4681–4689. https://doi.org/10.1109/BigData50022.2020.9378266

Muhamedyev, R., Yakunin, K., Iskakov, S., Sainova, S., Abdilmanova, A., & Kuchin, Y. (2015). Comparative analysis of classification algorithms. *2015 9th International Conference on Application of Information and Communication Technologies (AICT),* 96–101. https://doi.org/10.1109/ICAICT.2015.7338525

Novita Sari, N., Zain, I., Fithriasari, K., & Muhaimin, A. (2021, October 19). BR+ for Addressing Imbalanced Multi-label Data Classification Combined with Resampling Technique. https://doi.org/10.4108/eai.19-12-2020.2309179

Pan, Di., Zheng, X., Liu, W., Li, M., Ma, M., Zhou, Y., Yang, L., & Wang, P. (2020). Multi-label classification for clinical text with feature-level attention. *Proceedings - 2020 IEEE 6th Intl Conference on Big Data Security on Cloud, BigDataSecurity 2020, 2020 IEEE Intl Conference on High Performance and Smart Computing, HPSC 2020 and 2020 IEEE Intl Conference on Intelligent Data and Security, IDS 2020,* 186–191. https://doi.org/10.1109/BigDataSecurity-HPSC- IDS49724.2020.00042

Rahmawati, D., & Khodra, M. L. (2015). Automatic multi-label classification for Indonesian news articles. *2015 2nd International Conference on Advanced Informatics: Concepts, Theory, and Applications (ICAICTA)*, 1–6. https://doi.org/10.1109/ICAICTA.2015.7335382

Read, J., Puurula, A., & Bifet, A. (2014). Multi-label classification with meta-labels. Proceedings - *IEEE International Conference on Data Mining, ICDM, 2015-January(January),* 941–946. https://doi.org/10.1109/ICDM.2014.38

Sharma, S., & Mehrotra, D. (2018). Comparative analysis of multi-label classification algorithms. *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC),* 35–38. https://doi.org/10.1109/ICSCCC.2018.8703285

Tabia, K. (2019). Towards explainable multi-label classification. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, 2019-November,* 1088–1095. https://doi.org/10.1109/ICTAI.2019.00152

Taha, A. Y., Tiun, S., Abd Rahman, A. H., & Sabah, A. (2021). Multi-label over-sampling and under-sampling with class alignment for imbalanced multi-label text classification. *Journal of Information and Communication Technology, 20*(3), 423–456. https://doi.org/10.32890/jict2021.20.3.6

Venkatesan, R., & Er, M. J. (2014). Multi-label classification method based on extreme learning machines. *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV),* 619–624. https://doi.org/10.1109/ICARCV.2014.7064375

Wang, X., Liu, H., Yang, Z., Chu, J., Yao, L., Zhao, Z., & Zuo, B. (2017). Research and implementation of a multi-label learning algorithm for Chinese text classification. *Proceedings - 2017 3rd International Conference on Big Data Computing and Communications, BigCom 2017*, 68–76. https://doi.org/10.1109/BIGCOM.2017.34

Yee, T. W., Wild, C. J., & Yeet, T. W. (1996). Vector generalized additive models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(3), 481-493. https://www.jstor.org/stable/2345888