JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 221 - 230 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.1866
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

MOBILE LEGEND GAME PREDICTION USING MACHINE LEARNING REGRESSION METHOD

I Gede Wiarta Sena[1*], Andi W. R. Emanuel [2]
[1]Departement of Informatika, Institut Informatika Indonesia, Surabaya, Indonesia
[2]Department of Informatika, Universitas Atma Jaya Yogyakarta, Yogyakarta, Indonesia
email: *dedek@ikado.ac,id

Abstract: A research institute explains that with 83.7 million people using the Internet, Indonesia is among the top 20 internet users globally. Various individual or group activities require an internet network, one of which is playing games, for developments in the gaming sector, especially the MOBA (Massive Online Battle Arena) genre game, is being hotly discussed. There are various kinds of MOBA genre games, one of which is the Mobile Legends game. Many E-Sport Mobile Legends teams, especially in Asia, make this phenomenon a business space to generate large profits. In this study, the researcher recommends a good machine learning algorithm to predict the outcome of Mobile Legends matches. Of the 600 match history data analyzed, this study recommends the Artificial Neural Network (ANN) and Random Forest (RF) algorithms as the right algorithms to predict the outcome of the match. Prediction results from each algorithm can reach 82% and 80% accuracy. These findings can help the E-sports analysis team build their match strategy.

Keywords: artificial neural networ; machine learning; mobile legend; prediction; random forest

Abstrak: Sebuah lembaga penelitian menjelaskan bahwa dengan 83,7 juta penduduk yang menggunakan Internet, Indonesia termasuk di dalam 20 besar pengguna internet secara global. Berbagai aktivitas individu atau kelompok membutuhkan jaringan internet, salah satunya adalah bermain game, untuk perkembangan pada sektor game khususnya game bergenre MOBA (Massive Online Battle Arena) sedang hangat diperbincangkan. Ada berbagai macam game bergenre MOBA, salah satunya game Mobile Legends. Banyak tim E-Sport Mobile Legends khususnya di asia menjadikan fenomena ini sebagai ruang bisnis untuk menghasilkan keuntungnya yang besar. Dalam penelitian ini, peneliti merekomendasikan algoritma pembelajaran mesin yang baik untuk memprediksi hasil pertandingan Mobile Legends. Dari 600 data riwayat pertandingan yang dianalisis, penelitian ini merekomendasikan algoritma Artificial Neural Network (ANN) dan Random Forest (RF) sebagai algoritma yang tepat untuk memprediksi hasil pertandingan. Hasil prediksi dari masing-masing algoritma dapat mencapai 82% dan akurasi 80%. Temuan ini dapat membantu tim analisis E-sports membangun strategi pertandingan mereka.

Kata kunci: artificial neural network; machine learning; mobile legend; prediksi; random forest

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 221 - 230 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.1866
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

According to a research institute explained that of the 83.7 million people in the country's population, Indonesia is included in the top 20 internet users in the world [1]. With the development of time, this number will increase considering that accessing the internet is getting cheaper and easier. Various individual or group activities require an internet network, one of which is playing games [2]. The game is an interactive activation where one or more players follow the rules to complete the challenge to produce a final judgment [3][4]. So to the rampant developments in the game world today. The rapid development of games makes game developers compete to create quality games. In terms of developers, players (players) also take advantage of this situation to become proplayers in a game. Many gaming companies are competing to develop their e-sports teams for gaming tournaments.

One of the most popular games in the community is the MOBA genre game (Massive Online Battle Arena). MOBA is a type of game that combines two combinations of game genres RPG (Role Playing Game) and also RTS (Real-Time Strategy), where the player moves an in-game hero from two opposing teams to destroy the opponent's fortress [5][6]. Each hero played has different tasks, roles, strengths, and weaknesses, so a team must combine heroes from each player to achieve victory. Many MOBA games have been widely played, such as Mobile Legends, League of Legends, Vainglory, Dota 2, etc. From this problem, a process is needed to analyze match statistics that they can use to improve team performance. Strategy is needed to increase the quality of the e-sports team, with a quick and accurate analysis that coaches can use to improve the quality of players or strategies in the team. Using machine learning can produce useful information for predicting future events [7][8].

So, in this research, the process of predicting the victory of the Mobile Legends match will be carried out by utilizing the historical data of the previous match. The data used is in the form of match history, which is used as training data to get prediction results in the future. This research will recommend an accurate algorithm performance to make a prediction.

In the world of research, the prediction has become a widely discussed topic. With the development of today's technology, many researchers discuss predictions. This technique applies to many fields in society. In 2018 literature discussed predictions in the game world [9][10]. The utilization of machine learning technology is applied to predict the results of the NBA game, as seen from the statistical results of the previous game [10]. In this case, it can help coaches develop their strategies to benefit the next game. It was explained that the method used was the development of the Hybrid Fuzzy-SVM (HFSVM) model by combining the Fuzzy technique and also Support Vector Machines (SVM). The results of this study explain that the approach used obtains promising predictive results [11]. This research also raises a game that is the highest trend globally. Research that discusses the game entitled Dota 2 also uses machine learning techniques to predict. What is done in this study is to compare the performance of several machine learning algorithms such as naive Bayes, logistic regression, and decision trees. But the findings in

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 221 - 230 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.1866
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

this study cannot be separated from the playing ability of each player [12][13][14].

Various research fields use machine learning as a technology to support their strategies. There are also many uses of machine learning in the game world that are very useful for many people's lives. In this study, machine learning techniques are used to predict a company's bankruptcy in North America [15]. This study compares the predictions generated from machine learning techniques and traditional statistical techniques. It was explained that the machine learning technique was 10% more accurate than the traditional technique. Each result from the method used was 87% random forest, 69% logistic regression, and 50% linear discriminant analysis. The health sector also utilizes machine learning techniques for the necessary predictions. In this study, an analysis of medical data was conducted to predict the effectiveness of a chronic disease outbreak in a community that is frequently affected by the disease. This study involves data taken from a hospital in China in 2013 – 2015. This study uses a convolutional neural network (CNN) approach. The prediction accuracy generated using this approach reached 94.8% [16].

Mobile legends bang bang are the top 5 most popular games in the East Asian continent. There hasn't been much research that discusses this game. Machine learning is very helpful in analyzing data, one of which is to determine a good strategy in an E-sport Team Mobile Legends Bang Bang. Therefore, this study aims to provid recommendations for the most optimal machine learning algorithm used to predict victory. Researchers conducted trials on several machine learning algorithms and concluded the 2 best algorithms that could be used.

METHOD

Research methodology is an important part of the research process [17]. The flow of research will be carried out from the data collection process to the data analysis process to obtain research results. Image 1 is the flow of the methodology used in this study.

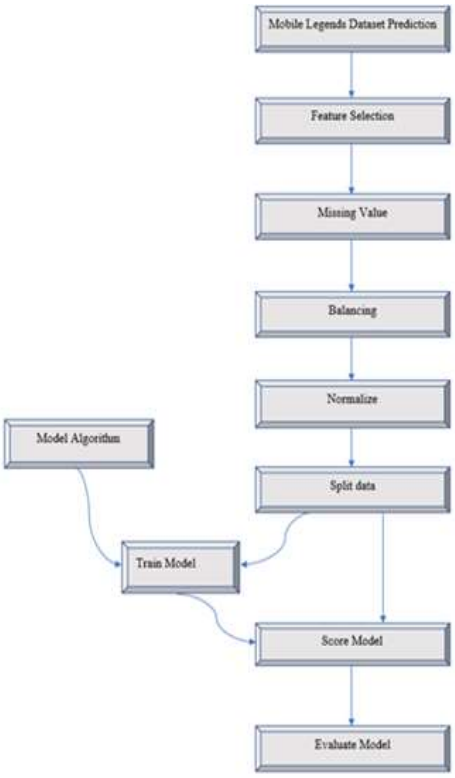

Image 1. Methodology

Feature selection

When data mining and machine learning are combined to achieve a goal, it is necessary to consider future selection to optimize the results of that goal [18][19]. The data used in this study is

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 221 - 230 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.1866
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

from historical statistics of match results used from previous matches. To further improve the performance of prediction accuracy, the match data used will be taken from the history of matches on the mythic tier. This tier is the highest in the Mobile Legends game. Statistical results from mobile legend matches reflect a team's strategy when competing.

There are 852 data collected from several Mobile Legends communities. The features used are average gold, average level, total kills, first blood, first turtle, first lord. The following table of features used in this study.

Table 1. Mobile Legends Feature

No	*Feature*	*Description*
1	avgGold	The average gold obtained by each team in the match
2	avgLvl	The average level obtained by each team in the match
3	ttlKill	The total number of kills obtained by each team in the match
4	frstBlood	Is a team that can kill the opponent the first time after the game starts
5	frstTurtle	Is the team that can kill the turtle the first time after the game starts
6	frstLord	Is a team that can kill the first lord after the game starts

Missing values

The data collection process is carried out at this stage, where the first dataset obtained is still not perfect [20][21]. This process removes empty datasets using the dropna function in python. Of the 852 data obtained after this process, the remaining dataset is only 832.

Balancing

In this stage, balancing is done on the output used in the dataset. In the data mining process there needs to be a process of balancing the number of data sets used based on data labels [22][23]. The next process will look for the maximum result among the five comparisons.The first comparison of the number of win labels will be applied by 60% and the lose label by 40%, for the next condition determines win label by 70% and the lose label by 30%, then the third comparison is 50% for the win label and 50% for the lose label, then 40% for the win label and 60% for the lose label, and the last 30% for the win label and 70% for the lose label.

Normalize

In this stage, data normalization is carried out on the dataset. Normalization makes the numbers in the dataset more structured [24][25]. Because the dataset used is not well structured, data normalization is needed. Normalize data using preprocessing scale function from python [26].

$$Xnew = ((Xold-Xmin))/((Xmax-Xmin)) \quad (1)$$

Description:
X = Dataset value
Xnew = The result of the value of X after preprocessing
Xold = Target value to be preprocessed
Xmin = The smallest value of all X values
Xmax = The largest value of all X values

Split data

Split data is very important for

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 221 - 230 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.1866
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

data analysis [27]. The next process divides the data frame into two categories, namely training data and testing data. The training data is used in the algorithm model as training data to achieve a prediction in the machine learning algorithm, while the testing data is used as test data to see accuracy, or in other words, to see the performance of the built model.

Model algoritm

A computer machine that can produce information automatically with the experiences provided will produce something extraordinary. The next process is designing several algorithm models to compare accuracy scores to recommend the right algorithm model in predicting mobile legends matches.

Table II. Model Algorithm

No	*Model algorithm machine learning*
1	*Artificial Neural network*
2	*Random forest*
3	*K-nearest Neighbor*
4	*Logistic Regression*
5	*Support Vector Machine*
6	*Decision Tree*

Evaluation

The evaluation stage is the last in this research. The next process uses several evaluation processes to determine the best machine learning algorithm model. The next process evaluates the findings from the design of the algorithm model through the confusion matrix table [28].

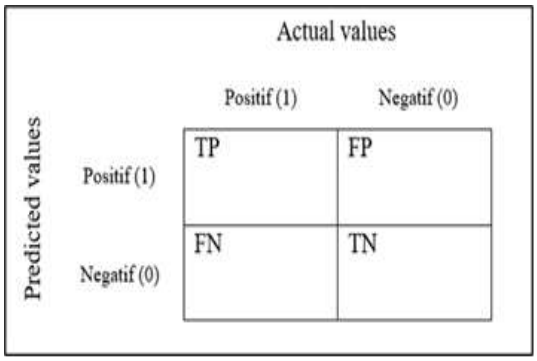

Image 2. Confusion matrix

RESULT AND DISCUSSION

The prediction of the Mobile legends bang bang (MLBB) matches results by utilizing historical data from previous matches and analyzing using several algorithm models. The two best algorithm models were obtained from the analysis results, namely Neural network and Random forest. And with the results of the Balancing process with the number of labels win 50% and label lose 50%, the maximum accuracy results are 82% for the neural network model and 80% for the random forest model. The following are the results of testing the six proposed algorithm models.

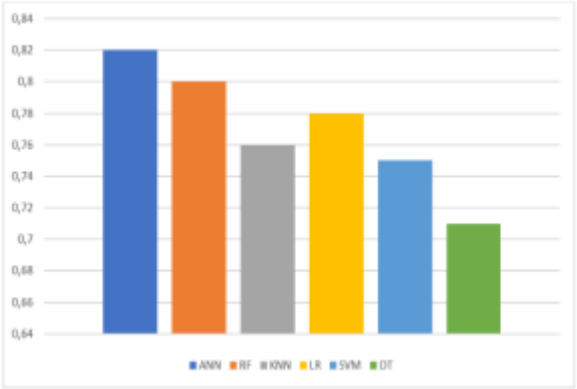

Image 3. Accuration report

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 221 - 230 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.1866
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Image 3 describes the accuracy scores of each algorithm model tested using a dataset made with features of selection, missing value, balancing, normalizing, and split data. Analysis of 6 models of machine learning algorithms was carried out, where the artificial neural network algorithm model got an accuracy score of 0.82, the random forest algorithm model with a score of 0.80, the k-nearest neighbor algorithm model got an accuracy score of 0.76, the logistic regression algorithm model got an accuracy score of 0.78, the support vector machine algorithm with an accuracy score of 0.75 and the last is a decision tree algorithm model with an accuracy score of 0.71. Based on the test results above, this study recommends two algorithm models with the highest accuracy scores: artificial neural network and random forest.

Confusion matrix and classification report evaluation

The evaluation results from the above model testing process will be displayed in this section. The accuracy generated by the artificial neural network algorithm model reaches 81.00%, and the random forest algorithm model accuracy reaches 80.00%. The details of the test results can be described in table 3, following the confusion matrix.

Table 3. Confusion matrix of ANN and RF

Prediction Value ANN	**Actual value**	
	Positive (*win*)	Negative (*lose*)
Positive class (*win*)	54	10
Negative class (*lose*)	13	43

Prediction Value RF	Actual value	
	Positive (*win*)	Negative (*lose*)
Positive class (*win*)	101	31
Negative class (*lose*)	17	91

In addition to the accuracy value, this study also displays precision, recall, and f1-score values obtained from the confusion matrix table. Precision shows the results of the level of accuracy between the requested data and the results predicted by the algorithm model, while recall serves to display the model's success in retrieving information, and the f1-score is the weighted average value of the results of precision and recall values. Classification report on the recommended algorithm model [29][30].

Table 4. Peformance report of ANN and RF

ANN	***Precision***	***Recall***	***F1-score***
Win	0.86	0.76	0.81
Lose	0.74	0.85	0.79
Avg	0.79	0.805	0.80

RF	***Precision***	***Recall***	***F1-score***
Win	0.81	0.84	0.82
Lose	0.81	0.77	0.79
Avg	0.81	0.805	0.805

Table 4 describes the precision, recall, and F1-score from the artificial neural network and random forest algorithm models. For the results of the ANN model, the precision value of the win label is 0.86, and the lose label is 0.74, so the average precision value is 0.79 or

79%. Then for the recall value of the win label, the result is 0.76, and the label lose 0.85, so the average result is 0.805 or 80.5%. And lastly, the f1-score value for the win label is 0.81, and the lose label is 0.79, so the average value of the f1-score obtained is 0.80 or 80%. Then for the results of the RF model, the precision value of the win label is 0.81, and the lose label is 0.81, so the average precision value is 0.81 or 81%. Then for the recall value of the win label, the result is 0.84, and the label lose 0.77, so the average result is 0.805 or 80.5%. And finally, the f1-score value for the win label is 0.82, and the lose label is 0.79, so the average value of the f1-score obtained is 0.805 or 80.5%.

CONCLUSION

In this study, from 600 datasets that have been preprocessed and tested with six proposed algorithm models, the study can conclude that the artificial neural network and random forest algorithm models are the most appropriate algorithm models for predicting the results of the Mobile Legends Bang Bang game. (MLBB) with 82.00% and 80.00% accuracy, respectively, among the four other algorithm models tested. Balancing normalizing can affect the accuracy of the artificial neural network and random forest algorithm models. In designing the artificial neural network algorithm model, the percentage of training and test data, the number of neurons in the hidden layer, and the number of epochs used can affect the model's score results. In this particular case, the researcher uses 80% of the training data and 20% of the test data from the entire dataset, using eight neurons in 1 hidden layer and 50 epochs. While designing the random forest algorithm model, it is necessary to pay attention to the number of nodes used. In this study are uses 100 nodes because it can produce the maximum accuracy score.

BIBLIOGRAPHY

[1] A. Katona, R. Spick, V. J. Hodge, S. Demediuk, F. Blok, A. Drachen, J. A. Walker, "Time to Die: Death Prediction in Dota 2 Using Deep Learning," *Conf. Proc. - IEEE Conference on Computatonal Intelligence and Games, CIGKatona, A., Spick*, 2019, doi: 10.1109/CIG.2019.8847997.

[2] A. S. Chan, F. Fachrizal, and A. R. Lubis, "Outcome Prediction Using Naïve Bayes Algorithm in the Selection of Role Hero Mobile Legend," *J. Phys. Conf. Ser.*, vol. 1566, no. 1, 2020, doi: 10.1088/1742-6596/1566/1/012041.

[3] U. Tokac, E. Novak, and C. G. Thompson, “Effects of game-based learning on students’ mathematics achievement: A meta-analysis,” J. Comput. Assist. Learn., vol. 35, no. 3, pp. 407–420, 2019, doi: 10.1111/jcal.12347.

[4] L. Dahabiyeh, M. S. Najjar, and D. Agrawal, “The effect of risk levels on technology adoption decision: the case of online games,” Inf. Technol. People, vol. 33, no. 5, pp. 1445–1464, 2020, doi: 10.1108/ITP-09-2019-0455.

[5] Z. Yu, M. Gao, and L. Wang, “The Effect of Educational Games

on Learning Outcomes, Student Motivation, Engagement and Satisfaction," J. Educ. Comput. Res., vol. 59, no. 3, pp. 522–546, 2021,doi:10.1177/0735633120969214.

[6] M. Mora-Cantallops and M. Á. Sicilia, "MOBA games: A literature review," *Entertain. Comput.*, vol. 26, no. February, pp. 128–138, 2018, doi: 10.1016/j.entcom.2018.02.005.

[7] J. M. Zhang, M. Harman, L. Ma, and Y. Liu, "Machine Learning Testing: Survey, Landscapes and Horizons," *IEEE Trans. Softw. Eng.*, pp. 1–1, 2020, doi: 10.1109/tse.2019.2962027.

[8] V. J. Hodge, S. Devlin, N. Sephton, F. Block, P. I. Cowling, and A. Drachen, "Win Prediction in Multiplayer Esports: Live Professional Match Prediction," *IEEE Trans. Games*, vol. 13, no. 4, pp. 368–379, 2021, doi: 10.1109/TG.2019.2948469.

[9] K. Passi and N. Pandey, "Increased Prediction Accuracy in the Game of Cricket Using Machine Learning," *Int. J. Data Min. Knowl. Manag. Process*, vol. 8, no. 2, pp. 19–36, 2018, doi: 10.5121/ijdkp.2018.8203.

[10] J. Le Louedec, T. Guntz, J. L. Crowley, and D. Vaufreydaz, "Deep learning investigation for chess player attention prediction using eye-tracking and game data," *Eye Track. Res. Appl. Symp.*, 2019, doi: 10.1145/3314111.3319827.

[11] V. J. Hodge, S. Devlin, N. Sephton, F. Block, P. I. Cowling, and A. Drachen, "Win Prediction in Multiplayer Esports: Live Professional Match Prediction," IEEE Trans. Games, vol. 13, no. 4, pp. 368–379, 2021, doi: 10.1109/TG.2019.2948469.

[12] D. Gourdeau and L. Archambault, "Discriminative Neural Network for Hero Selection in Professional Heroes of the Storm and DOTA 2," IEEE Trans. Games, vol. 13, no. 4, pp. 380–387, 2021, doi: 10.1109/TG.2020.2972463.

[13] M. Aung *et al.*, "Predicting Skill Learning in a Large, Longitudinal MOBA Dataset," *IEEE Conf. Comput. Intell. Games, CIG*, vol. 2018-August, pp. 1–7, 2018, doi: 10.1109/CIG.2018.8490431.

[14] K. Akhmedov and A. H. Phan, "Machine learning models for DOTA 2 outcomes prediction," pp. 1–11, 2021, [Online]. Available: http://arxiv.org/abs/2106.01782.

[15] J. P. Lai, Y. M. Chang, C. H. Chen, and P. F. Pai, "A survey of machine learning models in renewable energy predictions," Appl. Sci., vol. 10, no. 17, 2020, doi: 10.3390/app10175975.

[16] J. P. Lai, Y. M. Chang, C. H. Chen, and P. F. Pai, "A survey of machine learning models in renewable energy predictions," Appl. Sci., vol. 10, no. 17, 2020, doi: 10.3390/app10175975.

[17] W. Gu, K. Foster, J. Shang, and L. Wei, "A game-predicting expert system using big data and machine learning," *Expert Syst. Appl.*, vol. 130, pp. 293–305, 2019, doi: 10.1016/j.eswa.2019.04.025.

[18] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A Comprehensive Review of

Dimensionality Reduction Techniques for Feature Selection and Feature Extraction," J. Appl. Sci. Technol. Trends, vol. 1, no. 2, pp. 56–70, 2020, doi: 10.38094/jastt1224.

[19] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018, doi: 10.1016/j.neucom.2017.11.077.

[20] T. F. Johnson, N. J. B. Isaac, A. Paviolo, and M. González-Suárez, "Handling missing values in trait data," *Glob. Ecol. Biogeogr.*, vol. 30, no. 1, pp. 51–62, 2021, doi: 10.1111/geb.13185.

[21] S. A. Zahin, C. F. Ahmed, and T. Alam, "An effective method for classification with missing values," *Appl. Intell.*, vol. 48, no. 10, pp. 3209–3230, 2018, doi: 10.1007/s10489-018-1139-9.

[22] J. Kiani, C. Camp, and S. Pezeshk, "On the application of machine learning techniques to derive seismic fragility curves," *Comput. Struct.*, vol. 218, no. xxxx, pp. 108–122, 2019, doi: 10.1016/j.compstruc.2019.03.004.

[23] V. Romeo *et al.*, "Machine learning analysis of MRI-derived texture features to predict placenta accreta spectrum in patients with placenta previa," *Magn. Reson. Imaging*, vol. 64, pp. 71–76, 2019, doi: 10.1016/j.mri.2019.05.017.

[24] F. Jia, Y. Lei, N. Lu, and S. Xing, "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization," *Mech. Syst. Signal Process.*, vol. 110, pp. 349–367, 2018, doi: 10.1016/j.ymssp.2018.03.025.

[25] H. Yang *et al.*, "Immune-Related Prognostic Model in Colon Cancer: A Gene Expression-Based Study," *Front. Genet.*, vol. 11, no. May, pp. 1–10, 2020, doi: 10.3389/fgene.2020.00401.

[26] S. Gupta, R. Gupta, M. Ojha, and K. P. Singh, "A comparative analysis of various regularization techniques to solve overfitting problem in artificial neural network," *Commun. Comput. Inf. Sci.*, vol. 799, pp. 363–371, 2018, doi: 10.1007/978-981-10-8527-7_30.

[27] I. Santos, L. Castro, N. Rodriguez-Fernandez, Á. Torrente-Patiño, and A. Carballal, Artificial Neural Networks and Deep Learning in the Visual Arts: a review, vol. 33, no. 1. 2021. doi: 10.1007/s00521-020-05565-4.

[28] T. Wisanwanichthan and M. Thammawichai, "A Double-Layered Hybrid Approach for Network Intrusion Detection System Using Combined Naive Bayes and SVM," IEEE Access, vol. 9, pp. 138432–138450, 2021, doi: 10.1109/ACCESS.2021.3118573.

[29] W. Muhammad, M. Mushtaq, K. N. Junejo, and M. Y. Khan, "Sentiment analysis of product reviews in the absence of labelled data using supervised learning approaches," *Malaysian J. Comput. Sci.*, vol. 33, no. 2, pp. 118–132, 2020, doi: 10.22452/mjcs.vol33no2.3.

[30] M. Bouazizi and T. Ohtsuki, "Multi-class sentiment analysis on twitter: Classification

performance and challenges," *Big Data Min. Anal.*, vol. 2, no. 3, pp. 181–194, 2019, doi: 10.26599/BDMA.2019.9020002.