DOI https://doi.org/10.5281/zenodo.10049652#134

TWIST, 2024, Vol. 19, Issue 2, pp. 92-97

TWIST

Journal homepage: www.twistjournal.net

Machine Learning Approaches in Classifying Income Levels

Emely L. Laspiñas
College of Development Management, University of Southeastern Philippines – Mintal Campus,
Mintal, Davao City 8000, Philippines

John Vianne B. Murcia*
[1]- College of Development Management, University of Southeastern Philippines – Mintal Campus,
Mintal, Davao City 8000, Philippines

[2]- College of Business Administration Education, University of Mindanao,
Bolton Street, Davao City 8000, Philippines
[**Corresponding author*]

Abstract
This study compares the predictive accuracy of six machine learning classifiers – Logistic, Decision Tree (J48), RandomForest, Random Tree, IBk (k-NN), and NaiveBayes – for estimating adult income. Utilizing metrics such as true positive (TP) rate, false positive (FP) rate, precision, recall, and the F-measure, the performance of these classifiers was evaluated. Based on the results, RandomForest and Random Tree classifiers demonstrated the highest efficacy across all metrics. Nonetheless, other classifiers, such as Decision Tree and IBk, demonstrated promise, especially when the parameters were modified. The findings highlight the importance of model selection and fine-tuning in predictive modeling. These findings have significant ramifications for income forecasting, highlighting the capacity of machine learning to facilitate accurate socioeconomic forecasting. The study's results provide vital guidance for deploying the most appropriate classifier based on the specifics of the income dataset and the prediction task.

Keywords
Machine learning, Weka, Income levels, Classifiers, Classification

INTRODUCTION

The introduction of machine learning (ML) in economic analysis represents a significant change from traditional econometric methods to a more sophisticated, data-driven approach. This transformation is based on the acknowledgement of the inherent constraints of conventional models in comprehending the complex dynamics of income distribution. The classification and prediction of income levels have become crucial tasks due to the growing intricacy of global economic systems (Bandeira-Morais, Swart & Jordaan, 2021). Conventional models, which are bound by linear assumptions and a restricted capacity to analyze complex data, sometimes fall short in accurately or comprehensively explaining income differences (Yuan & Ling, 2020; Gelman, Hill & Vehtari, 2020). On the other hand, machine learning techniques provide a hopeful alternative that may effectively analyze large datasets and reveal intricate patterns that impact economic groups (Athey, 2018). This methodological development has repercussions that go beyond academic interest. It has major effects on policy making and socio-economic planning.

Econometric models have historically been fundamental to economic analysis, providing useful insights into the factors that contribute to differences in income. Nevertheless, Atkinson, Piketty, and Saez (2011) have highlighted that these models are hindered by their dependence on linear associations and a limited range of factors, which may obfuscate the complex dynamics of income distribution. This critique emphasizes a significant deficiency in conventional methods, specifically their incapacity to effectively handle the non-linear and diverse nature of economic data. The constraints of econometric models are most apparent when addressing the intricate interaction of variables that influence income levels, ranging from regional economic circumstances to worldwide market dynamics.

There is an increasing agreement among scholars that a new analytical framework is necessary in order to better understand and represent the actual economic inequalities. The incorporation of machine learning into the examination of income levels presents a revolutionary method that can effectively tackle the limitations of conventional econometric models. Delen, Oztekin, and Kong (2010) have showcased the capacity of machine learning algorithms to detect non-linear associations and previously undiscovered indicators of income, providing a more dynamic comprehension of economic inequalities. The ability to do detailed analysis is reinforced by the studies of Atkeson and Kehoe (2001) and Hastie, Tibshirani, and Friedman (2009), who highlight the capability of machine learning to effectively process intricate and multidimensional datasets. Through the utilization of machine learning techniques, researchers are able to surpass the limitations of traditional models, thereby gaining access to fresh perspectives on the variables that impact income distribution. This, in turn, establishes a stronger and more reliable basis for economic predictions.

The significance of machine learning in predicting income levels is emphasized by the increasing apprehension regarding income disparity at both the global and local levels. In his study, Milanovic (2016) illuminates the increasing disparity between the most affluent and least privileged communities worldwide, a phenomenon that presents substantial obstacles to both societal unity and long-term progress. In the Philippines, income disparities have a significant impact on economic growth, social mobility, and general societal well-being (Dacuycuy, 2019), making the situation particularly severe (Kelley & Evans, 2017). Moreover, scholars such as Ostry et al. (2014) and Beck, Demirgüç-Kunt, and Levine (2007) have emphasized the negative effects of excessive income disparity, emphasizing the importance of accurate, data-based analysis to guide policy decisions and foster fair growth.

The research disparity refers to the inadequate utilization of machine learning methods in predicting the income levels of adults. The objective of this study is to create a sophisticated predictive model that can accurately analyze and incorporate complicated income patterns. This will be achieved by combining global, national, and local data sources with advanced algorithms. Utilizing machine learning techniques can improve our understanding of the factors that determine income and increase the accuracy of income predictions. This study aims to rectify this insufficiency by contributing to the existing body of knowledge on income prediction through the utilization of machine learning techniques.

This study makes a substantial contribution to the field of economic research by utilizing machine learning techniques to provide a more comprehensive and precise analysis of income levels and inequalities. By utilizing the findings of previous studies (e.g., Alejandrino, Bolacoy, & Murcia, 2023; Biol & Murcia, 2024; Credit, 2022; Ngai, Xiu, & Chau, 2009), all who have demonstrated the enhanced prediction precision of machine learning (ML) models compared to conventional statistical approaches, this study not only contributes to academic knowledge but also offers a useful tool for policymakers. The utilization of machine learning to precisely categorize and forecast income levels presents opportunities for focused interventions, facilitating the efficient distribution of resources and the development of strategies to diminish income disparity and promote comprehensive economic expansion.

MATERIALS AND METHODS

Dataset

The Adult Income Prediction dataset (Patki, n.d.), which is readily accessible on the Kaggle platform, is a vast collection of demographic and employment-related information. Its primary objective is to facilitate research and forecasting of adult income levels. Due to its diverse set of characteristics, it provides an excellent basis for implementing and evaluating machine learning techniques in the context of income prediction. Given various socioeconomic and demographic factors derived from census data, the real-world nature of the data collection affords a unique opportunity to determine whether a person's annual income exceeds or falls below $50,000.

Table 1 List of Attributes of the Income Levels Prediction Dataset

Feature	Type	Description
Age	Numerical	Age of the individual
Work class	Categorical	Type of employment of the individual
Education	Categorical	Highest level of education achieved by the individual
Marital-status	Categorical	Marital status of the individual
Occupation	Categorical	Occupation of the individual
Relationship	Categorical	Relationship status of the individual
Race	Categorical	Race of the individual
Sex	Categorical	Sex of the individual
Capital-gain	Numerical	Capital gain of the individual
Capital-loss	Numerical	Capital loss of the individual
Hours-per-week	Numerical	Number of hours worked per week by the individual
Native-country	Categorical	Native country of the individual
Income	Categorical	Whether the individual's income is above or below $50,000 per year

The dataset has 16,281 observations or entries, each representing an individual. Each person is described by thirteen distinct characteristics or attributes, providing a multifaceted view of the population. These characteristics comprise various aspects of a person's existence, including age, occupation, level of education, and marital status. Variables of different types - numeric and categorical - increase the dataset's diversity, making it optimal for complex, real-world machine-learning applications.

The target variable, 'Income,' is a crucial component of the dataset. This binary categorical variable indicates whether an individual's annual income exceeds or falls below $50,000. This dichotomous classification serves as the outcome to be predicted based on the constellation of other features in the dataset and serves as the basis for predictive modeling efforts. This dataset is genuinely distinguished by its rich variety of features, which permits the training of nuanced machine-learning models for income forecasting. Due to the comprehensive nature of the data, machine learning models can use it to identify intricate relationships between demographic and employment-related factors. Therefore, it can provide a more comprehensive comprehension of their collective and individual effects on a person's income. This dataset offers a unique opportunity to delve deeper into the predictive analysis of income, aiming to improve the accuracy and reliability of such analyses in real-world scenarios.

Feature Selection

Using the Weka software for attribute selection, a systematic analysis of the Adult Income Prediction dataset's attributes was conducted based on their information gain and, consequently, their predictive power. The InfoGainAttributeEval evaluates a feature by calculating the information gain relative to a specific class. On the other hand, the Ranker search method assigns a rank to an attribute based on its assessed value (Hassan, & Khan, 2017). This combination of techniques assesses the potential information gain or reduction in entropy if the attribute in question were utilized for prediction. This evaluation determines the attribute's ability to distinguish data based on the class value, in this case, whether the income is greater than or less than $50,000.

Upon evaluation, the Ranker method ranks the attributes according to their scores, thereby listing attributes from most significant to least. As illustrated by our dataset, this classification revealed the 'Relationship' attribute to be the most predictive of income levels, with a value of 0.16575. This result is likely attributable to the socioeconomic factors associated with familial financial structures (Martin, 2006). In contrast, characteristics such as 'Race' and 'Native Country', which scored the lowest at 0.00819 and 0.00901 respectively, manifested less predictive power. The second and third-ranked attributes, 'Marital Status' and 'Capital Gain,' with scores of 0.15809 and 0.11337, respectively, supported previous research highlighting the significant economic implications of marital status (Schoeni, 1995) and the direct impact of a person's financial transactions and investments on their income (Frank, 2010). Moreover, occupation (0.09056), age (0.09521), education (0.08943), hours-per-week (0.05645), capital loss (0.04937), sex (0.03572), and work class (0.02423) were also ranked. This method emphasizes the significance of attribute selection in augmenting the effectiveness, interpretability, and precision of predictive models.

Data Classification and Cross-Validation

In the context of the Adult Income Prediction study, four prominent classifiers, including decision trees (J48) and RandomForest, *k*-nearest neighbors (*k*-NN) using Weka's IBk algorithm, and Naive Bayes, were utilized to classify the training data and predict the class label within the test set. In addition to these classifiers, Logistic Regression was employed due to its effectiveness in binary predictions, which corresponds to the binary character of income levels in this data set.

J48 decision trees and *k*-NN classifiers were tuned for their parameters. Adjustments were made to the confidence factor, which is essential for post-pruning efficacy. The investigation involved executing the J48 classifier with 0.25, 0.5, and 0.75 confidence values. For *k*-NN, the value of *k*, the number of nearest neighbors, was modified in order to decrease error probabilities (Dudani, 1976). The analysis was conducted with k values of 3, 5, 7, and 9 because it is widely accepted that the efficacy of *k*-NN improves ask increases (Alejandrino et al., 2023).

The classifier was evaluated using 10-fold cross-validation, a technique that guarantees a balanced and unbiased estimation of model performance (Kohavi, 1995). This method divides the data into ten subsets, uses nine subsets for training and one for testing iteratively, and then averages the performance measures.

RESULTS AND DISCUSSION

In this study, the classification performance of various machine learning classifiers was assessed utilizing the Adult Income Prediction dataset. Six classifiers, which includes Logistic, J48 (a decision tree classifier), Random Forest, Random Tree, IBk (representing *k*-NN), and NaiveBayes, were selected based on their widespread use and demonstrated efficacy in similar studies. To ensure a comprehensive evaluation of the model's predictive accuracy, a total of 11 classification tests were conducted, with results displayed in Table 2.

The evaluation of the classification precision of the selected classifiers in predicting the income level of the Adult Income Prediction dataset revealed that the models provided varying degrees of accuracy. The accuracy of Logistic was 85.82%, correctly classifying 13,923 instances. This result is consistent with Peng, Lee, and Ingersoll’s (2002) earlier finding, demonstrating that Logistic is an effective classifier for binary outcomes. Meanwhile, the accuracy of the J48 decision tree classifier improved as the confidence level increased when tested with three distinct confidence levels. Tests results revealed accuracy of 87.21%, 88.96%, and 90.84% for confidence levels of 0.25, 0.50, and 0.75, respectively. This result suggests that a higher confidence level may improve the efficacy of the classifier (Rajesh & Karthikeyan, 2017).

In addition, the table shows that the Random Forest classifier and the Random Tree classifier demonstrated the effectiveness of ensemble methods in improving prediction accuracy (Liaw & Wiener, 2002). The Random Forest classifier and the Random Tree classifier achieved accuracy levels of 98.35% and 98.37%, respectively. Similarly, the

IBk algorithm, which represents the k-NN method, was evaluated with k values of 3, 5, 7, and 9. It produced accuracy rates ranging from 89.11% (k=3) to 85.74 % (k=9), indicating an optimal range of k values for this particular dataset, in contrast to earlier research that suggested an increase in accuracy with higher k values (Alejandrino et al., 2023). Lastly, the accuracy of the Naive Bayes classifier was 82.24%. Despite having the lowest accuracy among the tested classifiers, Naive Bayes remains a popular option due to its simplicity and effectiveness with large data sets (John & Langley, 1995

Table 2 Classification accuracy of classifiers on the training dataset

Classifier	Variants	Correctly Predicted	Percentage
Logistic	-	13,923	85.82%
Decision tree (J48)	C 0.25	14,199	87.21%
	C 0.50	14,483	88.96%
	C 0.75	14,791	90.8482
Random Forest	-	16,013	98.35%
Random Tree	-	16,016	98.37%
Ibk (K-NN)	3	14,509	89.11%
	5	14,208	87.27%
	7	14,057	86.34%
	9	13,960	85.74%
NaiveBayes		13,389	82.24%

On the training data, the performance of six classifiers, including Logistic, Decision Tree (J48), RandomForest, Random Tree, and IBk (K-NN) and NaiveBayes was assessed. Their performance was evaluated based on the true positive rate, the false positive rate, precision, recall, and the F-measure. Table 3 displays the specific results of this performance evaluation. The results of the performance evaluation of the six classifiers provide significant insights into the behaviors and characteristics of various classification models, which have practical applications.

Table 3 Classification performance of the utilized classifiers on training data

Classifier	Variants	κ	TP rate	FP rate	Precision	Recall	F-measure
Logistic	-	0.5717	0..855	0.319	0.849	0.855	0.85
Decision Tree (J48)	0.25	0.6065	0.872	0.326	0.868	0.872	0.864
	0.50	0.6715	0.89	0.264	0.886	0.89	0.885
	0.75	0.433	0.821	0.443	0.808	0.821	0.806
Random Forest	-	0.9541	0.984	0.035	0.983	0.984	0.983
Random Tree	-	0.9541	0.984	0.047	0.984	0.984	0.984
lBk (k-NN)	3	0.6812	0.891	0.247	0.888	0.891	0.888
	5	0.6245	0.873	0.286	0.868	0.873	0.868
	7	0.5943	0.863	0.309	0.858	0.863	0.858
	9	0.5735	0.857	0.326	0.851	0.857	0.851
Naïve Bayes	-	0.4361	0.822	0.442	0.809	0.822	0.807

With F-measures of 0.983 and 0.984, respectively, the RandomForest and Random Tree classifiers have been observed to have the highest performance. These high scores suggest that these classifiers were able to establish a good balance between precision and recall, thereby reducing both the false-positive and false-negative rates. This is a strong indication that the models were able to generalize well from the training data and accurately predict the class labels of new, unobserved data.

The high performance of the RandomForest classifier may be attributed to the ensemble nature of this classifier, which is intended to enhance the overall result by minimizing overfitting. Similarly, the Random Tree classifier benefited from random feature selection, which contributed to its robust performance (Zhou et al., 2020). The robust performance of both classifiers suggests that they will likely be dependable for other classification tasks. In a variety of application domains where precision is of the utmost importance, their sturdiness makes them a secure option.

As indicated by the F-measure, the Decision Tree (J48) classifier exhibits a trend in which increasing the confidence factor (C) leads to decreased performance. This indicates that a larger confidence factor may result in overfitting, in which the model becomes overly complex and performs well on training data but unfavorably on unobserved data.

Looking on the results of the *k*-nearest neighbors (IBk) classifier, efficacy diminishes as neighborhood size (*k*) increases. This may suggest that larger *k* values incorporate more disturbance into the classification process, as more distant neighbors (who may belong to other classes) are considered (Franco-Lopez, Ek & Bauer, 2001). This suggests that a smaller neighborhood size may be preferable when the dataset contains classes with complex or irregular boundaries.

The average performance of the Logistic classifier may be attributable to the nature of logistic regression, which implies a linear decision boundary and may not perform well if the actual decision boundary is non-linear. Finally, despite its simplicity and the assumption of feature independence, the Naive Bayes classifier performs comparably to the J48 decision tree classifier with a confidence level of 0.75.

The results demonstrate that there is no universal classifier and that the selection of a classifier should be based on the task and dataset at hand. Moreover, significant number of scholars (e.g., Blachnik, 2017; Telikani et al., 2021; Yapp et al., 2020; Zhang et al., 2020) emphasize the significance of evaluating multiple classifiers and adjusting hyperparameters, such as the confidence factor in the Decision Tree classifier and the neighborhood size in the k-NN classifier, in order to find the model that best satisfies the task's particular requirements. Due to their inherent ability to prevent overfitting and consistently high performance (Aria, Cuccurullo & Gnasso, 2021; Kunapuli, 2023; Zhou, 2012), ensemble methods, such as RandomForest and Random Tree, may be an excellent starting point for many tasks.

CONCLUSION

This study concludes with crucial insights regarding the use of machine learning classifiers for estimating adult income levels. RandomForest and Random Tree classifiers demonstrated superior precision, recall, and F-measure performance, indicating their suitability for this task. Nevertheless, the performance of J48 Decision Tree and IBk classifiers, which could be altered by modifying parameters, demonstrates the adaptability that these models can offer. Even the simplicity of the NaiveBayes classifier can be useful for income level prediction, producing competitive results. Although Logistic algorithm's performance was average, it should not be undervalued for certain categories of income datasets. Ultimately, the selection of the optimal machine learning classifier for predicting adult income is dependent on the dataset and the task. These results highlight the significance of meticulous model selection and refining in the development of accurate and reliable adult income prediction models.

RECOMMENDATIONS

The current study highlights the profound capacity of machine learning (ML) methods to improve the accuracy of classifying income levels, which has important implications for economic research and policy development. The strong evidence produced by using classifiers like Random Forest and Random Tree, which are renowned for their high precision and capability to handle intricate data structures, indicates that policymakers and economic researchers should incorporate these advanced machine learning algorithms into their analytical frameworks. By engaging in this process, individuals can acquire a more profound comprehension of the complex mechanisms that dictate the allocation of income and detect subtle patterns that conventional econometric models may fail to recognize. This method not only simplifies the process of creating more focused and efficient socio-economic policies, but also assists in devising measures to reduce income inequalities. Therefore, harnessing the predictive capabilities of machine learning could have a crucial impact on promoting economic fairness and facilitating sustainable development.

This study presents various avenues that should be further investigated in future research. Firstly, it is necessary to examine the suitability of these machine learning classifiers in various economic settings and datasets, thus confirming their adaptability and reliability in forecasting income levels. Additionally, by fine-tuning the parameters of the classifiers, it is possible to achieve even greater levels of predicted accuracy. This implies that investing efforts in optimizing hyperparameters could lead to substantial benefits. Furthermore, incorporating supplementary socio-economic factors and implementing a more detailed method of classifying income could improve the models' ability to detect minor economic patterns and changes. Finally, integrating machine learning with knowledge from behavioral economics, sociology, and psychology can offer a comprehensive perspective on the factors that impact income levels. This can lead to the creation of detailed and sophisticated economic models. Participating in these research activities will not only progress the discipline of economic analysis but also aid in the development of policies that better tackle the intricate concerns of income inequality and economic diversity.

REFERENCES

1. Alejandrino, J. C., Bolacoy Jr, J., & Murcia, J. V. B. (2023). Supervised and unsupervised data mining approaches in loan default prediction. *International Journal of Electrical & Computer Engineering*, *13*(2), 1837-1847.
2. Aria, M., Cuccurullo, C., & Gnasso, A. (2021). A comparison among interpretative proposals for Random Forests. *Machine Learning with Applications*, *6*, 100094.
3. Athey, S. (2018). The impact of machine learning on economics. *The Economics of Artificial Intelligence: An Agenda*, 507-547.
4. Atkeson, A., & Kehoe, P. J. (2001). The transition to a new economy after the second industrial revolution. In *Federal Reserve Bank of San Francisco Proceedings*. National Bureau of Economic Research.
5. Atkinson, A. B., Piketty, T., & Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature*, *49*(1), 3-71.
6. Bandeira-Morais, M., Swart, J., & Jordaan, J. A. (2021). Economic complexity and inequality: does regional productive structure affect income inequality in Brazilian states*?*. *Sustainability*, *13*(2), 1006.
7. Beck, T., Demirgüç-Kunt, A., & Levine, R. (2007). Finance, inequality and the poor. *Journal of Economic Growth*, *12*, 27-49.
8. Biol, C. S., & Murcia, J. V. B. (2024). Supervised and Unsupervised Machine Learning Approaches in Predicting Startup Success. *TWIST*, *19*(1), 203-208.
9. Blachnik, M. (2017). Instance selection for classifier performance estimation in meta learning. *Entropy*, *19*(11), 583.
10. Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, *129*(4), 1553-1623.

11. Credit, K. (2022). Spatial models or random forest? Evaluating the use of spatially explicit machine learning methods to predict employment density around new transit stations in Los Angeles. *Geographical Analysis, 54*(1), 58-83.
12. Dacuycuy, L. B. (2019). Social mobility in the Philippine labor market. *Philippine Review of Economics, 55*(1&2), 1-40.
13. Dudani, S. A. (1978). The distance-weighted k-nearest neighbor rule. *IEEE Trans. on Systems, Man and Cybernetics, 8*(4), 311-313.
14. Franco-Lopez, H., Ek, A. R., & Bauer, M. E. (2001). Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sensing of Environment, 77*(3), 251-274.
15. Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and other stories. Cambridge University Press.
16. Hassan, S. A., & Khan, T. (2017). A machine learning model to predict the onset of Alzheimer disease using potential cerebrospinal fluid (CSF) biomarkers. *International Journal of Advanced Computer Science and Applications, 8*(12).
17. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: Springer.
18. John, G. H., & Langley, P. (2013). Estimating continuous distributions in Bayesian classifiers. arXiv preprint arXiv:1302.4964.
19. Kelley, J., & Evans, M. D. (2017). Societal inequality and individual subjective well-being: Results from 68 societies and over 200,000 individuals, 1981–2008. *Social Science Research, 62*, 1-23.
20. Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI* (Vol. 14, No. 2, pp. 1137-1145).
21. Kunapuli, G. (2023). Ensemble methods for machine learning. Simon and Schuster.
22. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18-22.
23. Martin, M. A. (2006). Family structure and income inequality in families with children, 1976 to 2000. *Demography, 43*(3), 421-445.
24. Mihaylova, B., Briggs, A., O'Hagan, A., & Thompson, S. G. (2011). Review of statistical methods for analysing healthcare resources and costs. *Health Economics, 20*(8), 897-916.
25. Milanovic, B. (2016). Global inequality: A new approach for the age of globalization. Harvard University Press.
26. Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications, 36*(2), 2592-2602.
27. Ostry, J. D., Berg, A., & Tsangarides, C. G. (2014). Redistribution, Inequality, and Growth, IMF Staff Discussion Note. International Monetary Fund.
28. Patki, S. (n.d.). Adult Income Prediction. Kaggle. Available at https://www.kaggle.com/datasets/saumitgp/adult-income-prediction.
29. Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*(1), 3-14.
30. Schoeni, R. F. (1995). Marital status and earnings in developed countries. *Journal of Population Economics, 8*(4), 351-359.
31. Telikani, A., Tahmassebi, A., Banzhaf, W., & Gandomi, A. H. (2021). Evolutionary machine learning: A survey. *ACM Computing Surveys (CSUR), 54*(8), 1-35.
32. Yapp, E. K., Li, X., Lu, W. F., & Tan, P. S. (2020). Comparison of base classifiers for multi-label learning. *Neurocomputing, 394*, 51-60.
33. Yuan, C. Z., & Ling, S. K. (2020, August). Long short-term memory model based agriculture commodity price prediction application. In *Proceedings of the 2020 2nd international conference on information technology and computer communications* (pp. 43-49).
34. Zhang, J., Wang, Y., Sun, Y., & Li, G. (2020). Strength of ensemble learning in multiclass classification of rockburst intensity. *International Journal for Numerical and Analytical Methods in Geomechanics, 44*(13), 1833-1853.
35. Zhou, Y., Cheng, G., Jiang, S., & Dai, M. (2020). Building an efficient intrusion detection system based on feature selection and ensemble classifier. *Computer Networks, 174*, 107247.
36. Zhou, Z. H. (2012). Ensemble methods: foundations and algorithms. CRC press.