
Implementing Machine Learning Techniques for Predicting Student Performance in an E-Learning Environment

Adi Suryaputra Paramita ^{1,*}, Laura Mahendratta Tjahjono ²

¹Information Systems Department, School of Information Technology Universitas Ciputra Surabaya, Indonesia

²Informatics Department, School of Information Technology Universitas Ciputra Surabaya, Indonesia

¹adi.suryaputra@ciputra.ac.id^{*}; ²laura@ciputra.ac.id;

* Corresponding author

(Received July 3, 2021 Revised August 11, 2021 Accepted August 29, 2021, Available online September 1, 2021)

Abstract

The pandemic of COVID-19 has altered the way people learn. Learning has moved from offline to online throughout this pandemic. Predicting student performance based on relevant data has opened up a new field for educational institutions to improve teaching and learning processes, as well as course curriculum adjustments. Machine learning technology can assist universities in forecasting student performance so that necessary changes in lecture delivery and curriculum can be made. The performance of the pupils was predicted using machine learning techniques in this research. Open University (OU) educational data is examined. Demographic, engagement, and performance metrics are used. The results of the experiment. The k-NN strategy outperformed all other algorithms on the OU dataset in some circumstances, but the ANN approach outperformed them all in others.

Keywords: E-Learning, Data Mining, Machine Learning, Student Performance

1. Introduction

The widespread usage of Internet technology has altered education from its conventional offline mode to an online/blended format known as the E-Learning Environment (ELE). For researchers, this has developed as a new area of study [1]. Jani et al. [2] reasoned that combining onsite learning with the ELE platform increased students' comprehension and performance. All academic institutions were closed and switched to online form during the COVID-19 epidemic, emphasizing the necessity of the E-learning Environment [3]. The most significant problem for educational institutions is accurate and reliable assessment of students' performance on ELE. It becomes extremely difficult and complex to e-access student performance without students cheating by using the Internet, printed notes, or other sources [4]. The prediction of real students' performance will be useful to teachers/course organizers in the early stages of the course when students require more attention and assistance [5].

Among all current methodologies, Educational Data Mining has been an excellent tool for identifying critical knowledge and patterns in massive educational datasets [6] over the last decade. It is the practice of pertaining data mining (DM) methods to data sets [7]. On the other hand, regression, classification as well as machine learning (ML), algorithms are more effective and accurate at forecasting student performance today. Prediction effectiveness and accuracy are vastly reliant on the data type of characteristics used, the dimension of the dataset, and the variety of the dataset. On the basis of regression and classification analysis of the ELE datasets, machine learning algorithms such as Decision Tree, SVC, k-NN, Random Forest, ANN, and AdaBoost are used to forecast students' performance.

2. Literature Review

The outcomes of assessing student performance with AI and Machine Learning are presented in this part. The forecast is based on teaching techniques, instructional materials, and data on access. The review is chronologically organized to show how the field has progressed through time. Elbadrawy constructed a class of linear multi-regression models for predicting student performance using educational data in a prior paper published in 2015. Former performance, participation with the Learning Management System (LMS), and course-related tasks were all

factored into the models. The purpose model was implemented with 11,556 student entries and 832 courses collected using a bespoke approach. The Root Mean Square Error (RMSE) of the multi regression model was 0.147, which was higher than the RMSE of the single regression model[8].

In 2016, Yee-King et al. [9] created a k-NN-based model for predicting collaborative social learning student success. To cope with the poor categorization, a multivariate classification method was utilized. In 2014, a custom-created dataset from a Coursera online course was used to validate the proposed method. The total number of User Interface (UI) clicks and mouse-overs generated over the course comprised the collected dataset. For the 2, 3, and 10 score bands, categorization accuracy was 88 percent, 77 percent, and 31 percent, respectively. Using k-NN and Support Vector Machine (SVM) machine learning techniques, Al-Shehri et al. [10] predicted students' final exam performance in 2017. A bespoke dataset from Portugal's University of Minho, which includes 395 data samples, was used to test the performance of machine learning models. Individual data variables and information about the students' families were included in the dataset. SVM is somewhat more accurate than k-NN, according to the research. Iqbal et al. [11] compared three distinct machine learning strategies for predicting student scores: CF, Matrix Factorization (MF), and Restricted Boltzmann Machines to validate machine learning algorithms (RBM). They used a proprietary dataset with 225 student records from Pakistan's International Technical University (ITU). Performance-based variables such as prior academic performance and interview score were used in the data collection process. With a root mean square error of 0.3, the RBM technique delivered the best results.

In 2018, Hussain et al. [12] used a number of learning-based algorithms to forecast student engagement and its impact on performance. To evaluate student involvement, they employed DT, Classification, Regression Tree (CART), JRIP Decision Rules, Gradient Boosting Trees (GBT), and Naive Bayes Classifier (NBC) on the OU dataset. For demographic, performance, and learning behavior analysis, only the July 2013 session (384 records) was used. The J48 decision tree approach was stated to surpass others with a maximum accuracy of 88.52 percent and a recall of 93.4 percent. During their initial review the same year, Heuer and Breiter [13] employed a range of machine learning algorithms to identify kids who were at risk. They looked at the OULAD Dataset, which included 32,593 student entries. The use of activity-based and performance topographies allowed for performance forecasting. They also used machine learning approaches like SVM, NB, RF, XGBoost, and Logistic Regression (LR), and found that SVM surpassed all other algorithms with an accuracy of 87.98%.

The application of machine learning approaches to predict and classify student performance was examined by Sekeroglu et al. [14] in 2019. Long-Short Term Memory, Backpropagation, and Support Vector Regression (SVR) were used for forecasting, and Backpropagation(BP), Support Vector Machine(SVM), and Gradient Boosting Classifier were used for classification (GBC). The Student Performance Dataset (SPD) was utilized for prediction analysis, while the Students' Academic Performance Dataset (SAPD) was used for classification analysis. The databases include information on students' demographic profiles, academic backgrounds, and behavioral pattern attributes. The authors claim that SVR is the best forecasting method and Backpropagation is the best classification algorithm. El Fouki et al. [15] then developed an upgraded classification model for predicting student performance using deep learning and Principal Component Analysis (PCA).

The proposed multi-dimensional technique decreases the dimensionality of the data and extracts critical information to increase the model's classification accuracy. A 496-record dataset that includes factors including student achievement, section information, and activity participation. After dimensionality reduction with PCA, a deep learning model, Multi-Layer Perceptron (MLP), and Bayes Net were employed to evaluate the data. With a classification accuracy of 92.54 percent, deep learning surpassed the other strategies. Hussain et al. [16] employed deep learning and the Adam optimizer to anticipate student achievement for the following year, based on internal assessment. Aside from the deep learning model, the Artificial Immune Recognition System (AIRS) v2.0 and AdaBoost were utilized to compare the results. The researchers used a proprietary dataset of 10,140 records from three Indian universities to conduct their investigation. The key component of the dataset used to predict final grades was students' performance on numerous exams. The best performance was a deep learning model with binary cross entropy loss and sigmoid activation, which had a classification accuracy of 95.34 percent. Later that year, Ajibade et al. [17] employed behavioural learning data and a range of classification algorithms to predict student progress. Differential Evolution (DE) was also employed to select behavioral features. The proposed methodologies were tested on a custom dataset of 500 student records. Demographic, academic, learning process, and behavioral learning characteristics were included in the dataset. The DT, k-NN, and SVM algorithms were all utilized, with the DT technique surpassing the others by a wide margin.

Tomasevic et al. [18] investigated the effect of numerous factors on student assessment predicting using Bayesian Linear Regression (BLR), SVM, k-NN, ANN, DT, Regularized Linear Regression (RLR), and statistical approaches in 2020. OULAD was studied in terms of demographics, engagement, and performance by the authors. The F1 score

and root mean square error (RMSE) were utilized as performance measures for classification and regression models. Based on engagement and performance features, the authors reported a 96.62 percent F1 score for an ANN and a 96.04 percent F1 score for an SVM (RBF kernel) based on demographic, engagement, and performance data. Hooshyar et al. [19] created a novel PPP technique based on tardiness to predict students' performance the next year. The presented technique significantly depended on assignment submission behavior to forecast a student's achievement. A custom dataset of 242 students from the Estonian University of Tartu was used to validate the suggested technique. Radial SVM (R-SVM), DT, Gaussian Process (GP), RF, Linear SVM (L-SVM), NN, AdaBoost, and NB were among the machine learning techniques employed to evaluate the dataset. Finally, the Neural Network had the greatest category feature accuracy of 96 percent, while the LSVM had the highest overall classification accuracy of 95 percent. Waheed also proposed using massive datasets from a VLE to anticipate student performance in the coming year using Deep Neural Networks (DNNs).

The researchers looked at the OULAD dataset, which contains 32,593 student records and is open-source. The dataset comprised demographic information, clickstream behavior, and assessment results. The proposed deep learning-based technique beat classical regression and SVM algorithms with an accuracy of up to 93 percent[20].

3. Methodology

3.1. Random Forest

A random forest is a classification and prediction technique in machine learning. It employs ensemble learning, a technique for resolving complex problems by combining multiple classifiers. A random forest algorithm is made up of many decision trees. The random forest technique trains a 'forest' using bagging or bootstrap aggregation. Bagging is a meta-algorithm that combines machine learning techniques to improve accuracy. The outcome is determined by the (random forest) algorithm using decision tree predictions. It forecasts by averaging or averaging the output of multiple trees. The output becomes more precise as the number of trees increases. The disadvantages of a decision tree algorithm are avoided by utilizing a random forest technique. It improves precision while decreasing the dataset's overfitting. It generates forecasts without the need for a large number of package shapes.

This is a method for supervised learning. A random forest method is created by combining multiple decision trees. In other words, it is a collection of numerous classification trees. It is utilized in the categorization and regression analysis processes. Each decision tree is framed by a set of rules. The decision tree algorithm will include specific rules for the provided training dataset with targets and features. Unlike decision trees, random forests do not require the user to calculate information gain in order to find root nodes. It calculates and saves the outcome of each randomly generated decision tree, as well as the vote for each anticipated objective. In this case, the random forest algorithm's final forecast is the one with the most votes. For additional information, see [20].

3.2. Naive-Bayes

The Naive Bayes method is a simple but productive predictive modelling approach. The model contains two distinct types of probabilities that can be calculated easily using the training data: (1) the likelihood of each class and (2) the conditional probability for each class given each x value. After calculating the probability model, the Bayes theorem can be used to make predictions for new data. When dealing with real-valued data, it is common to assume a Gaussian distribution in order to make estimation of these probabilities easier (bell curve). The term "naive Bayes" refers to the assumption that each input variable is independent. Despite the fact that this is a significant assumption that is implausible when dealing with real-world data, the technique works well in a variety of complex situations.

Hussain explains: Naive Bayes is a Bayesian graphical model with nodes for each column or feature (NB)[12]. It is referred to as naive because it disregards prior parameter distributions and makes the assumption that all features and rows are independent. Ignoring the past has a number of benefits and drawbacks. We can apply any distribution to individual attributes and infer the most probable ones from the data, which is a significant advantage. To simplify probability and prior algebra, we do not have to restrict the class of prior distributions to the exponential family. It has the disadvantage of being a maximum likelihood model. The posterior does not improve iteratively. Regardless of its benefits and drawbacks, the NB method remains a probabilistic generative model, which means it can generate data given parameters. These nodes generate values that correspond to observable feature values. Numeric attribute values can be a discrete collection of symbols, whereas categorical attribute values can be a discrete set of symbols. The label column is used to indicate the location of a node. The label may be categorical or real-valued (as in a regression problem) (as in a classification problem).

It is based on the Bayes theorem, a classification technique. It determines the probability that an object will possess certain properties associated with a given class. As a result, it is occasionally referred to as a probabilistic classifier. The occurrence of one attribute is unrelated to the occurrence of other characteristics in this technique.

3.3. k-Nearest Neighbour (k-NN)

The k-nearest neighbors (KNN) technique is a data categorization technique that estimates the likelihood that a data point belongs to one of two categories based on the data points closest to it. The supervised machine learning algorithm k-nearest neighbor is used to solve classification and regression problems. However, it is primarily used to address categorization issues. KNN is a non-parametric, slow learning algorithm. Because it does not perform any training when you supply the training data, it is referred to as a lazy learning algorithm or lazy learner. Rather than that, it simply records the data and does not perform any calculations during the training phase. It does not begin modeling until the dataset is searched. As a result, KNN is an exceptional tool for data mining.

Fix et al. introduced k-NN as a non-parametric machine learning method in 1951. It categorizes input according to how well it is received in the immediate vicinity. It is unknown whether this is the optimal method for data dissemination. KNN generates predictions directly from the training set. As a result, it is the simplest and most appropriate method for generating predictions. For additional information, please see [22].

3.4. Support Vector Machine (SVM)

A support vector machine (SVM) is a supervised machine learning model that solves two-group classification problems by utilizing classification techniques. SVM models can categorize new text after being fed sets of labeled training data for each category. They have two significant advantages over more recent algorithms such as neural networks: they are faster and require less data (in the thousands). This makes the method particularly well-suited to text classification problems, which frequently require only a few thousand tagged examples.

Vapnik introduced SVM as a relatively robust supervised machine learning approach in 1963, which was later extended by Boser et al. [23]. This algorithm generates a large number of hyperplanes in a high-dimensional space with the goal of achieving good separation between them. The high margin between hyperplanes explains why there is little generalization loss. The test sample is divided into two groups. Each test sample is represented as an m-dimensional vector divided into two parts by a (m-1)-dimensional hyperplane. Multiple hyperplanes can be used to separate test samples, but in the linear classification situation, the one with the greatest separation is chosen [24,25].

3.5. Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) is a data processing technique or strategy inspired by the biological nervous system's operations, particularly in the processing of information by human brain cells. This technique relies heavily on the structure of the information processing system, which is unique and varies depending on the application. A neural network is a collection of interconnected information processing elements (neurons) that work together to solve a problem, most commonly a classification or prediction problem.

The Neural Network's functioning can be compared to supervised learning, which is how humans learn by using examples. Neural Networks are programmed to do certain tasks, such as pattern recognition or data classification, and then tweaked over time. Learning in biological systems includes changing existing synaptic connections between neurons; in the case of a Neural Network, this is done by updating the weight values in each input, neuron, and output link. An artificial neural network (ANN) is a machine learning system that simulates human brain behavior. It consists of nodes. The network is made up of layers, nodes, and connections. Nodes are used to represent artificial neurons. It possesses the ability to understand and relay input impulses to other neurons. In most circumstances, an ANN consists of input, output, and a number of hidden layers, each of which contains artificial neurons that are linked to one another. For more information, see [26,27].

4. Result and Discussion

This part describes the data set and procedures utilized for the experimental study, as well as the findings of the developed Naive Bayes, Random Forest, k-NN, SVM, and ANN Data Mining and Machine Learning algorithms for analyzing student performance based on various combinations of input features.

4.1. Dataset

The Open University dataset [28] was downloaded from Kaggle and utilized in both tests. There are 32,593 student records in this dataset from 15 different nations. Furthermore, the information includes student-chosen courses, demographic data, and student interactions with the e-learning environment. The dataset was cleaned, and the desirable features were extracted. For classification analysis, dataset cleaning entails dealing with missing values and assigning arithmetic values to phrases. The dataset's input features are demographic (D), engagement (E), and performance (P), with students' performance measured as pass or fail as the target variable.

4.2. Data Pre-Processing

A comma-separated file contains the final data set. Demographic, Engagement, and Past Performance are the three categories. The following are the features of the dataset:

Table 1. Details of Data set features

Demographic features	Values	Description
Gender	[0,1]	
Highest Education	0	0: Below high school
	0.25	0.25: High school
	0.50	0.5: Diploma
	0.75	0.75: Bachelor
	1	1: Post graduate
Age	0	0: <35
	0.5	0.5: 35-55
	1	1: >55
Engagement		
Total clicks	[0-N]	0 - N
Performance		
Score per assessment	[0-1]	0 - 100
No. of Attempts	[0-1]	0 – N
Final exam score	[0-1]	0 – N

Table 2. Performance comparison of Classification Algorithms to predict final exam result

		D	E	P	D+E	D+P	E+P	D+E+P
Accuracy	CNN	0.7591	0.9620	0.9884	0.9526	0.9934	0.9965	0.9922
	KNN	0.6905	0.9622	0.9992	0.9626	0.9966	0.9966	0.9934
	ANN	0.7594	0.9984	0.9990	0.9298	0.9982	0.9984	0.9979
	SVM	0.7594	0.9516	0.9984	0.9340	0.9963	0.9983	0.9958
	Naive Bayes	0.6804	0.9362	0.9893	0.9521	0.9865	0.9964	0.9945
	Random Forest	0.6805	0.9164	0.9970	0.9533	0.9782	0.9964	0.9947

		D	E	P	D+E	D+P	E+P	D+E+P
F1 Score	CNN	0.8531	0.9460	0.9979	0.9565	0.9886	0.9928	0.9944
	KNN	0.8108	0.9756	0.9995	0.9759	0.9977	0.9977	0.9956
	ANN	0.8633	0.9480	0.9989	0.9535	0.9987	0.9988	0.9986
	SVM	0.8632	0.9685	0.9993	0.9564	0.9975	0.9988	0.9972
	Naive Bayes	0.8412	0.9585	0.9983	0.9484	0.9778	0.9988	0.9911
	Random Forest	0.8436	0.9655	0.9773	0.9425	0.9855	0.9988	0.9862

		D	E	P	D+E	D+P	E+P	D+E+P
J-Index	CNN	0.6799	0.9521	0.9981	0.9500	0.9855	0.9855	0.9923
	KNN	0.6818	0.9524	0.9991	0.9530	0.9955	0.9955	0.9913
	ANN	0.7596	0.9012	0.9979	0.9112	0.9975	0.9977	0.9973
	SVM	0.7593	0.9290	0.9987	0.9165	0.9951	0.9977	0.9945
	Naive Bayes	0.7583	0.9270	0.9897	0.9055	0.9981	0.9937	0.9965
	Random Forest	0.7553	0.9370	0.9917	0.9075	0.9971	0.9957	0.9955

From the OU education dataset [28], algorithms were used to classify performance based on demographics, engagement, and historical performance. Table-2 shows the results of the experiment. Table-2 demonstrates that for the E, P, and D+E situations, k-NN performed best, with accuracy of 0.9622, 0.9992, and 0.9626, respectively. However, ANN fared best in the D, D+P, E+P, and D+E+P situations, with accuracy of 0.7594, 0.9982, 0.9984, and 0.9979, respectively.

5. Conclusion

Using multiple data mining and machine learning techniques, this article projected the dataset's performance. In the trials, data was cleaned and prepared in a CSV file, and then data mining and machine learning algorithms were used to predict student performance in final exams. The findings were presented in tabular style, with the best performing algorithms highlighted in the table. For various feature variations, the experimental findings reveal that the K-NN algorithm outperforms ANN and SVM, Naive Bayes, and Random Forests, with ANN outperforming other algorithms in some circumstances. In the future, more parameters will be measured using a real dataset, precision and recall will be calculated, and work on missing values will be expanded.

References

- [1] Sweta S. (2021) Educational Data Mining in E-Learning System. In: Modern Approach to Educational Data Mining and Its Applications. SpringerBriefs in Applied Sciences and Technology. Springer, Singapore. https://doi.org/10.1007/978-981-33-4681-9_1

- [2] J. Jani, R. Muszali, S. Nathan, and M. S. Abdullah, "Blended learning approach using frog vle platform towards students' achievement in teaching games for understanding," *Journal of Applied and Fundamental Sciences*, vol. 10, pp. 1131–1151, Jan. 2018.
- [3] I Chatziralli, C. V. Ventura, S. Touhami, R. Reynolds, M. Nassisi, T. Weinberg, K. PakzadVaezi, D. Anaya, M. Mustapha, A. Plant, M. Yuan, and A. Loewenstein, "Transforming ophthalmic education into virtual learning during COVID-19 pandemic: a global perspective," *Eye*, pp. 1–8, Jul. 2020, publisher: Nature Publishing Group.
- [4] H. Mellar, R. Peytcheva-Forsyth, S. Kochar, A. Karadeniz, and B. Yovkova, "Addressing cheating in e-assessment using student authentication and authorship checking systems: Teachers' perspectives," *International Journal for Educational Integrity*, vol. 14, p. 2, Feb. 2018.
- [5] C. Vegega, P. Pytel, and M. F. Pollo-Cattaneo, "Application of the Requirements Elicitation Process for the Construction of Intelligent System-Based Predictive Models in the Education Area," in *Applied Informatics, ser. Communications in Computer and Information Science*, H. Florez, M. Leon, J. M. Diaz-Nafria, and S. Belli, Eds. Cham:Springer International Publishing, 2019, pp. 43–58.
- [6] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," *Procedia Computer Science*, vol. 72, pp. 414–422, Jan. 2015
- [7] A. S. J. Abu Hammad, "Mining Educational Data to Analyze Students' Performance (A Case with University College of Science and Technology Students)," *Central European Researchers Journal*, vol. 4, no. 2, 2018.
- [8] A. Elbadrawy, R. Studham, and G. Karypis, "Personalized Multi-Regression Models for Predicting Students' Performance in Course Activities," Mar. 2015.
- [9] M. Yee-King, A. Grimalt-Reynes, and M. d'Inverno, "Predicting student grades from online, collaborative social learning metrics using K-NN." in *EDM, 2016*, pp. 654–655.
- [10] H. Al-Shehri, A. Al-Qarni, L. Al-Saati, A. Batoaq, H. Badukhen, S. Alrashed, J. Alhiyafi, and S. O. Olatunji, "Student performance prediction using support vector machine and k nearest neighbor," in *2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE), 2017*, pp. 1–4, tex.organization: IEEE.
- [11] Z. Iqbal, J. Qadir, A. N. Mian, and F. Kamiran, "Machine learning based student grade prediction: A case study," *arXiv preprint arXiv:1708.08744*, 2017.
- [12] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, "Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores," Oct. 2018, iSSN: 1687-5265 Pages: e6347186 Publisher: Hindawi Volume: 2018. [Online]. Available: <https://www.hindawi.com/journals/cin/2018/6347186/>
- [13] H. Heuer and A. Breiter, "Student success prediction and the tradeoff between big data and data minimization," *DeLFI 2018-Die 16. E-Learning Fachtagung Informatik, 2018*, publisher: Gesellschaft für Informatik eV.
- [14] B. Sekeroglu, K. Dimililer, and K. Tuncal, "Student Performance Prediction and Classification Using Machine Learning Algorithms," Mar. 2019, pp. 7–11.
- [15] M. El Fouki, N. Akin, and K. E. El Kadiri, "Multidimensional approach based on deep learning to improve the prediction performance of DNN models," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, no. 02, pp. 30–41, 2019.
- [16] S. Hussain, Z. Muhsen, Y. Salal, P. Theodorou, F. Kurtoglu, and G. Hazarika, "Prediction Model on Student Performance based on Internal Assessment using Deep Learning," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, p. 4, Apr. 2019.
- [17] S. Ajibade, N. Ahmad, and S. M. Shamsuddin, "Educational Data Mining: Enhancement of Student Performance model using Ensemble Methods," *IOP Conference Series: Materials Science and Engineering*, vol. 551, p. 012061, Aug. 2019.
- [18] N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," *Computers & Education*, vol. 143, p. 103676, Jan. 2020.
- [19] D. Hooshyar, M. Pedaste, Y. Yang, L. Malva, G.-J. Hwang, M. Wang, H. Lim, and D. Delev, "From Gaming to Computational Thinking: An Adaptive Educational Computer Game-Based Learning Approach," *Journal of Educational Computing Research*, p. 0735633120965919, Oct. 2020, publisher: SAGE Publications

- [20] Han, Jiawei, and Micheline Kamber. *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers, 2001.
- [21] E. Fix, *Discriminatory analysis: nonparametric discrimination, consistency properties*. USAF School of Aviation Medicine, 1951.
- [22] D. Coomans and D. L. Massart, "Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-nearest neighbour classification by using alternative voting rules," *Analytica Chimica Acta*, vol. 136, pp. 15–27, 1982.
- [23] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144–152.
- [24] Y. Pu, D. B. Apel, and H. Xu, "Rockburst prediction in kimberlite with unsupervised learning method and support vector classifier," *Tunnelling and Underground Space Technology*, vol. 90, pp. 12–18, 2019.
- [25] D. Fradkin and I. Muchnik, "Support vector machines for classification," *DIMACS series in discrete mathematics and theoretical computer science*, vol. 70, pp. 13–20, 2006.
- [26] A. Abraham, "Artificial neural networks," *Handbook of measuring system design*, 2005.
- [27] K. Mehrotra, C. K. Mohan, and S. Ranka, *Elements of artificial neural networks*. MIT press, 1997.
- [28] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open university learning analytics dataset," *Scientific data*, vol. 4, p. 170171, 2017.