www.mazedan.com

Available online at
www.mdl.mazedan.com

MIJAM
www.mazedan.com/mijam

THE DEVELOPMENT OF A MODEL OF AUGMENTED ENSEMBLE LEARNING TO ENHANCE STUDENT PERFORMANCE IN BEHAVIOUR ANALYSIS

AKASH A RAIPURE[1], YOGENDRA SHARMA[2], DYUTI BANERJEE[3], SADIK ABDUL HAMID KHAN[4]

MAZEDAN INT. JOURNAL OF APPLIED MATHEMATICS

e-ISSN:
Article id-MIJAM0102004
Vol.-1, Issue-2
Received: 10 Sep 2021
Revised: 11 Nov 2021
Accepted: 15 Nov 2021

Citation: Raipure, A. A., Sharma, Y., Banerjee, D. & Khan, S. A., H. (2021). The development of a model of augmented ensemble learning to enhance student performance in behaviour analysis. *Mazedan International Journal of Applied Mathematics*, 1(2), 19-24.

Abstract

Deep learning techniques classify acquired data into 1 of "N" behavioural groups. Pattern analysis and post-processing are used by these models. Their accuracy for certain behavioural classes restricts their therapeutic use. This book presents an improved ensemble learning methodology to scale student behaviour analysis. The proposed approach intelligently blends different deep learning architectures to identify student behaviour. Models include VGGNet-19, ResNet101, Inception Net, and Xception Net. VGGNet-19 and ResNet101 are used for short-range data sequences, whereas Inception Net and Xception Net are utilised for intermediate to long sequences. Long-range sequences include academic profiles, everyday activities, study routines, etc. An ensemble learning layer combines these models and evaluates student behaviour characteristics. The suggested model classifies input data with over 90% accuracy, 85% precision, and 89% recall, which is greater than typical behaviour analysis models. The "Student Life" collection includes user information, sensing data, educational information, survey data, and educational maintenance allowance (EMA) data from over 10000 students of varied ages and qualifications. The suggested augmented ensemble model is scalable and deployable because it achieves consistent performance for each entity. The suggested model beats most current models in terms of precision, recall, accuracy, and other performance criteria.

Keywords: Student, behaviour, clinical, machine learning, augmented, ensemble.

1. INTRODUCTION

Models for analysing student behaviour seek to perform analysis of multidomain data, which necessitates efficient data collection, pre-processing, feature extraction, feature selection, classification, and post-processing operations on the part of the models. These activities need the building of highly effective machine learning models that are able to perform pattern analysis on several datasets and then link those datasets using filtering and rule mining techniques. Figure 1 depicts a typical model for the analysis of student behaviour, in which several elements of a student's life, including as personal data, attendance records, library borrowing, test scores, passing % [14], etc. are analysed. This information is then sent to a layer known as an extraction transform and load (ETL) layer, which is responsible for transforming the data into features before it is used in subsequent processing. A feature selection unit is provided with these features, and it then combines those features with statistical and cluster-based analysis. In order to have an approximate idea of the classes that students are in, the results of the clustering are mapped against their labels. In order to create each of the aforementioned blocks, a broad number of machine learning models have been presented; these models differ from one another in terms of their accuracy, precision, recall, and computational delay performance. In the next paragraph, this article will provide an overview of these algorithms, as well as a discussion of their intricacies, benefits, limits, and downsides. As a result of this research, it has come to our attention that the majority of these models have high performance on a particular dataset; nevertheless, they are not scalable because of this. Following the review portion is a discussion about the suggested design of an enhanced ensemble learning model for increasing student behaviour analysis performance. This discussion follows immediately after the review section in order to increase this scalability. The model is tested on a variety of datasets, and its results are compared to those of a number of other models that are considered to be state-of-the-art. Lastly, this article draws to a close with some thought-provoking remarks on the suggested model, as well as some suggestions for how to enhance it.

[1]Department of Computer Science & Engineering, G H Raisoni University, Amravati, Maharashtra, India.
[2]Vishwakarma Institute of Information Technology, Pune – 411048, Maharashtra (India)
[3]NIMS University Jaipur, Shobha Nagar, Jaipur, Rajasthan 303121
[4]Department of ENT, College of Engineering & Technology North Maharashtra Knowledge City, Bambhori Pr. Chandsar, Maharashtra 425002, India
***Corresponding author email-** rais.khan@ghru.edu.in

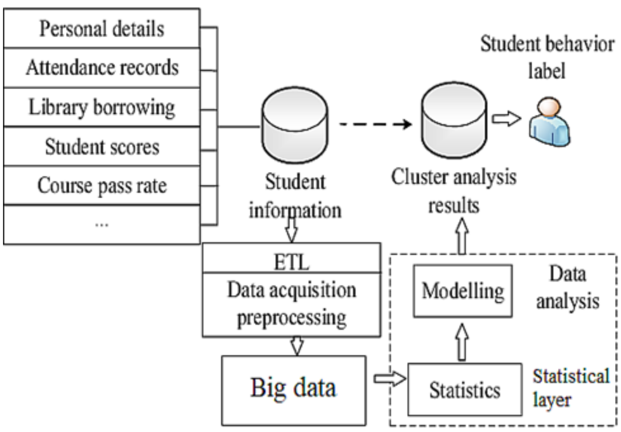

Figure 1 A typical student behaviour analysis model

2. LITERATURE REVIEW

A wide variety of system models are proposed for student analysis of student behaviour, with each of them utilising a different facet of student behaviour in order to arrive at an accurate forecast. For instance, the research presented in [1, 2] presents models for learner community analysis as well as highly accurate mathematical learning accomplishment analysis. For the purpose of carrying out this activity, these models make use of deep learning architectures, in addition to models that are efficient at feature extraction and selection. In a similar vein, the research presented in [3, 4, 5, 6] suggests the use of scalable operating systems, student trust and satisfaction levels, density-based clustering, as well as teaching quality feedback, gender, and the social and economic profile of the students in order to analyse their behaviour. The research presented in [7] suggests an unique multiple features fusion approach in conjunction with deep knowledge tracing for the purpose of producing very accurate predictions of student behaviour. The model has extremely high performance, in addition to low delay rates for the categorization and analysis of student behaviour.

In a similar fashion, visualisation models [8], models [9], eye movement analysis [10], and learning management system interface [11] are all methods that are capable of being utilised for the study of student behaviour in an effective manner. In addition, it has been discovered that green usage patterns [12], call patterns [13], examination performance analysis [14], and cyber bullying analysis [15] are all highly effective methods for analysing student behaviour. [12], [13], [14], and [15] In order to accomplish classification tasks, these models make use of several deep learning architectures, such as VGGNet, Q-Learning, deep nets, and so on. However, these models are only tested on one particular kind of student behaviour, which severely restricts their scalability as well as their deployment potential. In the next part, we will present an innovative ensemble learning model that is capable of supporting many datasets in order to eliminate this disadvantage.

3. DESIGN OF THE PROPOSED AUGMENTED ENSEMBLE LEARNING MODEL

From the literature review, it is observed that most of the work done in the field of student behaviour analysis uses context-sensitive information, and thus have limited applicability. Thus, the proposed augmented ensemble learning model [9] for student behaviour analysis [11] uses a flexible architecture which can be used for analysis of multiple types of datasets, thereby resulting in analysis of multiple student behavioural parameters. The model initially uses a data cleaning & pre-processing layer, wherein multiple sources of data are passed through outlier detection & data adaptation. This step ensures that all the data is available in a standard format for consecutive layers. The pre-processed data is given to an augmented combination of multiple deep learning architectures including VGGNet-19[15], ResNet101, Inception Net and Xception Net. Results of these architectures are combined using an ensemble layer, which assists in final classification. Once new behavioural classes are added to the system, it is retrained in order to incorporate them, and improve accuracy & scalability of the proposed model. Architecture of the proposed model is visualized in figure 2, wherein internal computational steps are showcased. It is observed that results of all the CNN architectures are given to an ensemble model for final student behaviour class prediction [11], which assists in improving overall process accuracy.

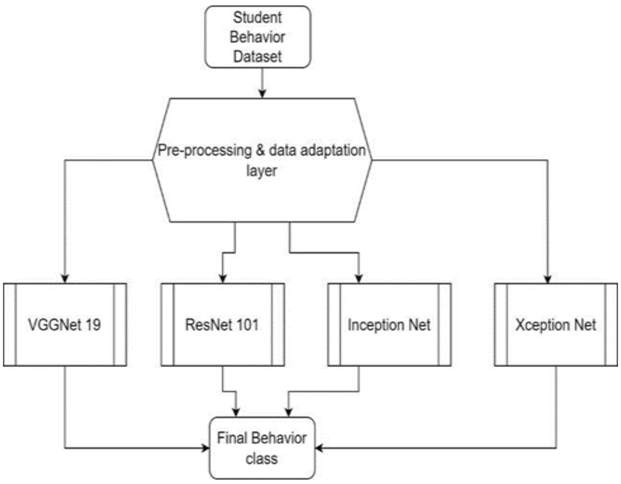

Figure 2 Architecture for the proposed model

The proposed model uses a pre-processing layer, which follows the given steps,

- All input data is converted into an array of bytes, and stored in different arrays.
- These arrays are separated w.r.t. parameter type, thus, all data belonging to student marks is stored in one array, while data belonging to their buying patterns [12] [13] is stored in another array.

This data is normalized using the following equation 1,

$$N = \frac{A_d - \min(A_d)}{\max(A_d) - \min(A_d)} \quad (1)$$

Where, Nd, and Ad represents normalized & actual data values.

- These data values are given to a pre- processing layer, which performs outlier removal using the following sub-process,
  - the normalized data is clustered using kMeans
  - for kMeans, initialize k=2, and perform clustering

- Evaluate intra-cluster distance using equation 2, where, internal values represent each instance inside the cluster,

$$D_{intra} \frac{\sum_{i=1}^{N} \sqrt{(x_i-x_j)^2-(y_i-y_j)^2+\cdots+(z_i-z_j)^2}}{N^2} \quad (2)$$

- Evaluation of intra-cluster distance is done using equation 2, where centroid values of each cluster are used.
- After this, sum of squared distances is used for cluster selection as per equation 3,

$$D_{inter} \frac{\sum_{i=1}^{N} D_{intra}+\sum_{i=1}^{N} D_{inter}}{N} \quad (3)$$

- These values are evaluated for k=2, 3, 4, 5, …, 20, and plotted against „k" as shown in figure 3,

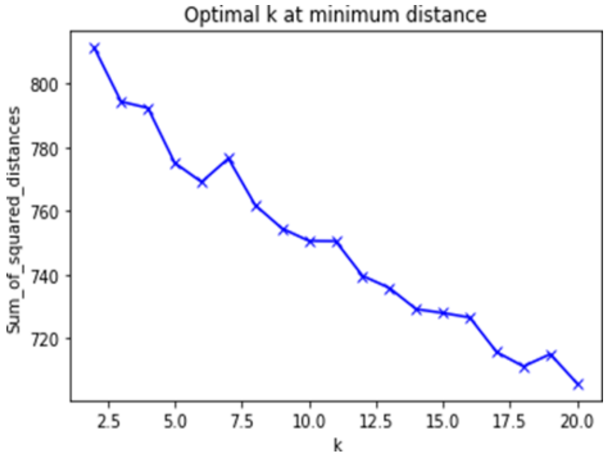

Figure 3 Sum of squared distances for different clustering configurations

- Value of „k", with minimum value of ds is selected for final clustering.

Based on this value of „k", cluster with minimum number of samples termed as outlier cluster, and its elements are removed from the dataset. After this, the values are given to an augmentation layer for behaviour classification. The architectures for each of these models is displayed in figure 4, 5, 6, and 7, wherein VGG-19, ResNet 101, Inception Net, and Xception Net are described

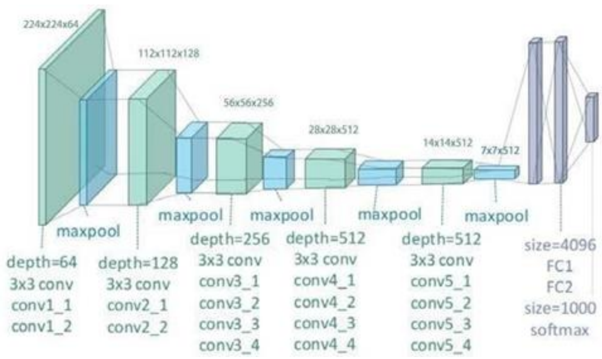

Figure 4 Architecture for the VGGNet-19 model

The VGGNet-19 [15] and ResNet101 models are used for short-ranged data sequences, which include datasets for eCommerce transactions, or

social media activity. The architecture for ResNet 101 model is described in figure 5 as follows,

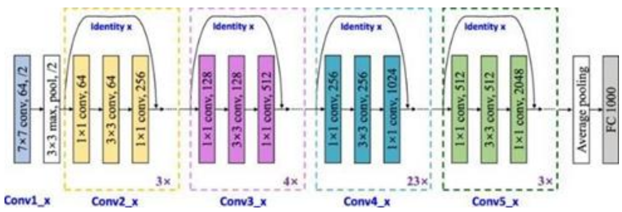

Figure 5 Architecture for the ResNet 101 model

It uses a combination of residual layers, which learn from each other in order to improve internal classification accuracy of the CNN model. This is followed by use of Inception Net model, which is described in figure 6 as follows,

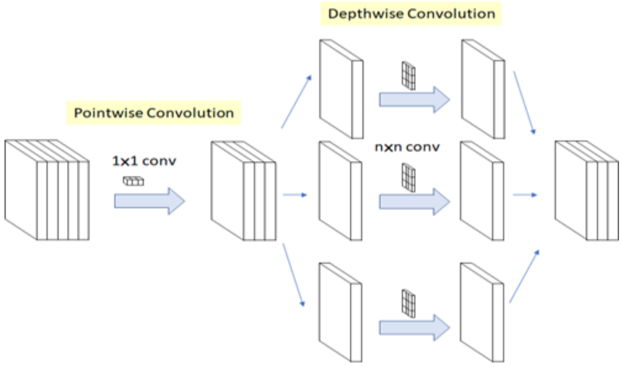

Figure 6 Model for pointwise and depth wise convolution

Classification results from each of these models are given to an ensemble learning layer, which works using the following steps,

- The processed sets are given to the following classifiers, and the indices of correctly classified instances (C) are tracked for each classifier,
  - VGGNet-19 ($VGGc$)
  - Res Net 101 model ($RESc$)
  - Inception Net ($Incepc$)
  - Xception Net model (X$cepc$)
- These instances are combined, and their unique behavioral classes are evaluated using equation 4,

$$C_{final} = Unique\ (\cup\ VGG_c, RES_c, Incep_c, Xcep_c) \quad (4)$$

- Final accuracy is evaluated via comparison of these classes with the testing dataset.
- For new inputs, feature sets are matched with input data patterns, and most probable classifier is selected for matching.

Based on this model, results are evaluated for precision, recall, accuracy, and computational delay. These results are compared with standard models, and are showcased in the next section.

4. RESULTS & COMPARISON

The proposed model is evaluated with [5], and [R2] in terms of accuracy, precision, recall, and delay measurements. The standard „Student Life" dataset is used for this purpose, and is available at https://studentlife.cs.dartmouth.edu/dataset.html, and can be used with open-source licensing. The dataset was taken for over 10000 students, and performance measures were evaluated. For instance, table 1 showcases the values of accuracy(A) for different number of students (NS), and different algorithms,

Table 1 Accuracy evaluation for different algorithms on different test set sizes

NS	A (%) [5]	A (%) [R2]	A (%) [Prop.]
1000	71.55	56.34	88.45
1500	72.22	56.86	89.27
2000	72.36	56.98	89.45
2500	72.66	57.21	89.82
3000	72.81	57.32	90.00
3500	72.89	57.38	90.10
4000	72.89	57.38	90.10
4500	72.90	57.39	90.11
5000	72.90	57.40	90.12
5500	72.90	57.40	90.12
6000	72.91	57.40	90.13
6500	72.91	57.40	90.13
7000	72.91	57.40	90.13
8500	72.92	57.41	90.14
10000	72.92	57.41	90.14

From the accuracy values it can be observed that the proposed model is 18% more efficient than existing implementations, which makes it useful for high accuracy applications. Similar observations are made for precision (P) values, and can be observed from the following table 2,

Table 2 Average precision values for different algorithms

NS	P (%) [5]	P (%) [R2]	P (%) [Prop.]
1000	68.80	63.61	85.03
1500	69.43	64.19	85.83
2000	69.58	64.32	86.01
2500	69.86	64.60	86.35
3000	70.01	64.73	86.53
3500	70.08	64.79	86.62
4000	70.08	64.79	86.62
4500	70.08	64.81	86.63
5000	70.10	64.81	86.65
5500	70.10	64.81	86.65
6000	70.10	64.82	86.65
6500	70.10	64.82	86.65
7000	70.10	64.82	86.65
8500	70.11	64.82	86.66
10000	70.11	64.82	86.66

From the precision values it can be observed that the proposed model is 16% more efficient than existing implementations, which makes it useful for high precision behaviour analysis applications. Similar observations are made for recall (R) values, and can be observed from the following table 3,

Table 3 Average recall values for different algorithms

NS	R (%) [5]	R (%) [R2]	R (%) [Prop.]
1000	65.33	60.40	80.75
1500	65.93	60.96	81.50
2000	66.06	61.09	81.66
2500	66.34	61.34	82.00
3000	66.46	61.46	82.16
3500	66.54	61.53	82.25
4000	66.54	61.53	82.25
4500	66.55	61.54	82.26
5000	66.55	61.54	82.28
5500	66.55	61.54	82.28
6000	66.56	61.55	82.28
6500	66.56	61.55	82.28
7000	66.56	61.55	82.28
8500	66.58	61.55	82.29
10000	66.58	61.55	82.29

From the recall values it can be observed that the proposed model is 15% more efficient than existing implementations, which makes it useful for high recall applications. Similar observationsare made for computational delay values, and can be observed from the following table 4,

Table 4 Average computational delay values for different algorithms

NS	Delay (ms) [5]	Delay (ms) [R2]	Delay (ms) [Prop.]
1000	12.61	16.02	10.20
1500	12.49	15.87	10.11
2000	12.47	15.84	10.09
2500	12.42	15.77	10.05
3000	12.39	15.74	10.03
3500	12.38	15.72	10.01
4000	12.38	15.72	10.01
4500	12.38	15.72	10.01
5000	12.38	15.72	10.01
5500	12.38	15.72	10.01
6000	12.37	15.72	10.01
6500	12.37	15.72	10.01
7000	12.37	15.72	10.01
8500	12.37	15.72	10.01
10000	12.37	15.72	10.01

From the delay values it can be observed that the proposed model is 20% more efficient than existing implementations, which makes it useful for high accuracy & high-speed applications. Due to high accuracy, high precision, high recall, and low delay values the underlying system model is highly efficient for classification of different student behavioural classification applications.

5. CONCLUSION & FUTURE SCOPE

From the result evaluation, it is observed that the proposed model is 18% more accurate than existing implementations, 16% more precise than existing implementations, has 15% more recall than existing implementations, and has 20% more speed than other models. This makes the proposedmodel highly applicable for real-time usage, and due to its high accuracy performance, it can be used for multiple behavioural class types, thereby further enhancing its deployment capabilities. In future, this work can be extended using Q-learning, and reinforcement learning mechanisms, which assist in obtaining better efficiency, with reduced delay requirements, thereby making it useful for real-time applications.

Deep learning techniques classify acquired data into 1 of "N" behavioural groups. Pattern analysis and post-processing are used by these models. Their accuracy for certain behavioural classes restricts their therapeutic use. This book presents an improved ensemble learning methodology to scale student behaviour analysis. The proposed approach intelligently blends different deep learning architectures to identify student behaviour. Models include VGGNet-19, ResNet101, Inception Net, and Xception Net. VGGNet-19 and ResNet101 are used for short-range data sequences, whereas Inception Net and Xception Net are utilised for intermediate to long sequences. Long-range sequences include academic profiles, everyday activities, study routines, etc. An ensemble learning layer combines these models and evaluates student behaviour characteristics. The suggested model classifies input data with over 90% accuracy, 85% precision, and 89% recall, which is greater than typical behaviour analysis models. The "Student Life" collection includes user information, sensing data, educational information, survey data, and educational maintenance allowance (EMA) data from over 10000 students of varied ages and qualifications. The suggested augmented ensemble model is scalable and deployable because it achieves consistent performance for each entity. The suggested model beats most current models in terms of precision, recall, accuracy, and other performance criteria.

REFERENCES

[1] Li, Chao et al. "Learner2Vec-based learner community evolution analysis–A case study involving student card data". IEEE access: practical innovations, open solutions 7 (2019): 27416–27425. Web.

[2] Hwang, Wu-Yuin et al. "Effect of ubiquitous fraction app on mathematics learning achievements and learning behaviors of Taiwanese students in authentic contexts". IEEE transactions on learning technologies 13.3 (2020): 530–539. Web.

[3] Wang, Lei et al. "A scalable operating system experiment platform supporting learning behavior analysis". IEEE transactions on education 63.3 (2020): 232–239. Web.

[4] Ismail, Mustapha, Erbug Celebi, en Halil Nadiri. "How student information system influence students‟ trust and satisfaction towards the university? An empirical study in a multicultural environment". IEEE access: practical innovations, open solutions 7 (2019): 111778–111789. Web.

[5] Li, Xiaoyong et al. "An unsupervised ensemble clustering approach for the analysis of student behavioral patterns". IEEE access: practical innovations, open solutions 9 (2021): 7076–7091. Web.

[6] Gupta, Anika, Deepak Garg, en Parteek Kumar. "Analysis of students‟ ratings of teaching quality to understand the role of gender and Socio-economic diversity in higher education". IEEE transactions on education 61.4 (2018): 319–327. Web.

[7] Liu, Dong et al. "Multiple features fusion attention mechanism enhanced deep knowledge tracing for student performance prediction". IEEE access: practical innovations, open solutions 8 (2020): 194894–194903. Web.

[8] Rosenthal, Michele. Mediating Religion, Sanctifying Media. Boston, MA: Walter de Gruyter, 2023. Print. Religion and Society.

[9] Dame, Christof et al., reds. Atlas of the Cellular and Molecular Development of Human Hematopoiesis. 1st ed. Berlin, Germany: Springer, 2023. Print.

[10] Burak, Sabahat, en Oğuzhan Atabek. "Yaratıcılık Özyeterliliği ve Tasarım Özyeterliliğinin Müziksel Davranışlar ve Kişisel Değişkenler Ile Olan İlişkileri". Kalem Uluslararasi Egitim ve Insan Bilimleri Dergisi 13.1/24 (2023): 35–57.

[11] Web.Demir, Okay, en Mustafa Demir. "Sınıf Öğretmenlerinin Bireysel Yenilikçilik Özellikleri Ile Öğrenmeye İlişkin Tutumları Arasındaki İlişkinin Belirlenmesi". Kalem Uluslararasi Egitim ve Insan Bilimleri Dergisi 13.1/24 (2023): 83–104. Web.

[12] Açıkel, Ayşenur, en Damla Yıldız. "Bartın Üniversitesi Ulus Meslek Yüksekokulu Öğrencilerinin Okula Aidiyet Duygularında DERs Dışı Etkinliklerin Yeri ve Önemi". Kalem Uluslararasi Egitim ve Insan Bilimleri Dergisi 13.1/24 (2023): 1–33. Web.

[13] Uygun, Kamil. "Social Studies Teacher Candidates Perceptions on Local Children s Games". Kalem Uluslararasi Egitim ve Insan Bilimleri Dergisi 13.1/24 (2023): 59–81. Web.

[14] Kawazoe, Yoshiyuki, en T. Kaneko. Compounds of Rare Earth Elements with Main Group Elements. 1st ed. Berlin, Germany: Springer, 2023. Print. Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology - New Series.

[15] Atakan, Baris, en Sebastià Galmés. AutonomousCommunication and Coordination Inspired by Nature. 1st ed. Basel, Switzerland: Springer International Publishing, 2023. Print. SpringerBriefs in Applied Sciences and Technology.

[16] Veronica Wawira, Gitonga, Ernest Saina, en Koech Ronald. "Factors Affecting Household Income among Motorcycle Operators. A Case of Embu County Residence, Kenya". Journal of business management and economic research 5.2 (2023): 48–56. Web.

[17] Costa, Breno et al. "Orchestration in Fog Computing: A Comprehensive Survey". ACM computing surveys 55.2 (2023): 1–34. Web.

[18] Adeleke, Oluwamayowa Ade, Nicholas Bastin, en Deniz Gurkan. "Network Traffic Generation: A Survey and Methodology". ACM computing surveys 55.2 (2023): 1–23. Web.

[19] Lai, Zhao-Rong, en Haisheng Yang. "A Survey on Gaps between Mean-Variance Approach and Exponential Growth Rate Approach for Portfolio Optimization". ACM computing surveys 55.2 (2023): 1–36. Web.

[20] Xie, Yiqun, Shashi Shekhar, en Yan Li. "Statistically-Robust Clustering Techniques for Mapping Spatial Hotspots: A Survey". ACM computing surveys 55.2 (2023): 1–38. Web.

[21] Sarwar, Tabinda et al. "The Secondary Use of Electronic Health Records for Data Mining: Data Characteristics and Challenges". ACM computing surveys 55.2 (2023): 1–40. Web.

[22] Zhou, Jingya et al. "Network Representation Learning: From Preprocessing, Feature Extraction to Node Embedding". ACM computing surveys 55.2 (2023): 1–35. Web.

[23] Hasan, Omar, Lionel Brunie, en Elisa Bertino. "Privacy-Preserving Reputation Systems Based on Blockchain and Other Cryptographic Building Blocks: A Survey". ACM computing surveys 55.2 (2023): 1–37. Web.

[24] Filipi Gonçalves dos Santos, Claudio et al. "Gait Recognition Based on Deep Learning: A Survey". ACM computing surveys 55.2 (2023): 1–34. Web.

[25] Jin, Qiao et al. "Biomedical Question Answering: A Survey of Approaches and Challenges". ACM computing surveys 55.2 (2023): 1–36. Web.

[26] Xia, Meng et al. "QLens: Visual analytics of multi-step problem-solving behaviors for improving question design". IEEE transactions on visualization and computer graphics 27.2 (2021): 870–880. Web.

[27] Xu, Zhuojia, Hua Yuan, en Qishan Liu. "Student performance prediction based on blended learning". IEEE transactions on education 64.1 (2021): 66–73. Web.

[28] Villa, A. Menacho, C. Perez Molina, en Manuel Castro. "Students‟ behavior when connecting to the LMS: A case study at UNED". IEEE Revista Iberoamericana de Tecnologias del Aprendizaje 14.3 (2019): 87–94. Web.

[29] Lin, Weijane et al. "Modeling reading behaviors: An automatic approach to eye movement analytics". IEEE access: practical innovations, open solutions 9 (2021): 63580–63590. Web.

[30] Tang, Hua et al. "Predicting green consumption behaviors of students using efficient firefly grey wolf-assisted K-nearest neighbor classifiers". IEEE access: practical innovations, open solutions 8 (2020): 35546–35562. Web.

[31] Vhaduri, Sudip, Sayanton Vhaduri Dibbo, en Yugyeong Kim. "Deriving college students‟ phone call patterns to improve student life". IEEE access: practical innovations, open solutions 9 (2021): 96453–96465. Web.

[32] Saba, Tanzila et al. "Categorizing the students‟ activities for automated exam proctoring using proposed deep L2-GraftNet CNN network and ASO based feature selection approach". IEEE access: practical innovations, open solutions 9 (2021): 47639– 47656. Web.

[33] Alotaibi, Norah Basheer. "Cyber bullying and the expected consequences on the students‟ academic achievement". IEEE access: practical innovations, open solutions 7 (2019): 153417–153431. Web.