ACADEMIC PERFORMANCE PREDICTION IN UNIVERSITIES USING ENSEMBLE ALGORITHMS: A LITERATURE REVIEW

INGRID CAROLINA LÓPEZ MESA¹, LEONARDO EMIRO CONTRERAS BRAVO² & HECTOR
JAVIER FUENTES LOPEZ³

¹Industrial engineering. Faculty of Engineering. Engineering Mater student. Universidad Distrital Francisco José de Caldas, Colombia.

²Mechanical Engineering. Faculty of Engineering. Universidad Distrital Francisco José de Caldas. Bogotá, Colombia.

³Economist. Faculty of Engineering. Universidad Distrital Francisco José de Caldas. Bogotá, Colombia.

ABSTRACT

One of the main challenges faced by higher education institutions is offering services with quality that are influenced by different variables such as the academic performance of their students. Due to the complexity and variety of factors that affect this parameter, national and international research projects have been carried out to analyze said factors including the intrinsic characteristics of students and external elements related to the environment and social relationships. This document details a literature review regarding machine learning techniques, particularly ensemble algorithms, for the early prediction of academic performance. The discussed factors are classified into academic, socio-demographic, psychosocial and online learning, and the most commonly used ensemble algorithms in academic performance prediction are identified.

KEYWORDS: Machine learning, higher education, ensemble methods, academic performance

Received: Oct 06, 2020; **Accepted:** Oct 26, 2020; **Published:** Dec 01, 2020; **Paper Id.:** IJMPERDOCT202080

1. INTRODUCTION

Given the complexity and variety of factors that interfere in academic performance, research projects have been carried out both nationally and internationally to analyze said factors, including the intrinsic characteristics of students up to external aspects regarding the environment and social behavior. (García-Tinizaray et al., 2019) argues that academic performance is the main indicator of the success (or failure) of a student which directly affects the quality and effectiveness of educational processes in higher education.

There are other indicators to assess educational processes such as academic retention, dropout and abandonment rate (Salcedo, 2010), the early prediction of academic performance and the variables affecting the student's environment, can mitigate the previously described metrics (García et al., 2014). The environment can be permeated by a variety of variables, which has led different research studies to determine which are most impactful. The analyzed variables have been often pedagogic and personal (García et al., 2014), sociodemographic, motivational, cognoscitive and emotional (Moreno & Chauta, 2012).

Machine learning (ML) algorithms in the student context have been proposed to process large amounts of data rendered available by institutions so that researchers can predict academic performance (Miguéis et al., 2018). Over the last years, studies on the prediction of academic performance have gained the interest of higher education

institutions, with the purpose of extracting common factors that surround this phenomenon and set actions to mitigate it. In this sense, numerous studies have been published over the last decade in which different machine learning techniques have been used to predict academic performance of students. These methods are presented in the subsequent section.

2. MATERIALS Y METHODS

Some definitions and concepts are clarified before approaching the methodology and results derived from this project.

2.1 Definitions of Academic Performance

One of the main academic challenges faced in higher education institutions consists on reducing poor academic performance of their students. Some authors such as (Caballero & Palacio, 2007) state that academic performance is the achievement of goals established in the curriculum. On the other hand, according to (Velázquez & Soriano, 2006), knowledge is an area measured directly by academic average scores compared to a minimum standard.

Among the reviewed definitions, most authors indicate that academic performance of an individual is mainly measured by the grades obtained in the enrolled subjects, which is reflected on the GPA, used by many universities to establish an assessment scale for students (Bunce & Hutchinson, 1993). In more recent studies, (York et al., 2015) state that academic success is achieved through the achievement of academic goals, career satisfaction, persistence, skills, individual skills and professional success. (Tian & Sun, 2018) refers to academic performance as the appropriation of knowledge that students reach by receiving guidance from teachers through curricular learning within a given time period.

2.2 Ensemble Algorithms

Nowadays, Machine Learning (ML) is defined as the science that uses algorithms to deliver something interesting based on datasets without typing any specific code to solve a problem (Rojas, 2008), that has many application fields (Figure 1). This area can be classified in different ways depending on how the system learns to develop, with the most common being Supervised Learning and Unsupervised Learning. In this field, algorithms capture information based on datasets and learn from said data, using a variety of models and methods that generate estimations of the studied problem whether it is programmed by humans or determined by the system using context (Russo et al., 2016).

The field of classical machine learning is a very important subgroup of ML. It uses two groups of tools known as supervised learning and unsupervised learning. The ensemble techniques offer an alternative solution through the aggregation of predictions obtained with supervised and unsupervised learning methods to generate a new prediction that is more reliable and accurate (Bourel, 2012). Some of the ensemble methods include:

Figure 1: Application of tasks in ML. Source: Authors

- **Bagging:** This algorithm divides a dataset into various subsets, taking elements randomly from the initial dataset. The same learning algorithm is executed in all data subsets and the final prediction is the combination obtained from a voting process including all answers (Bourel, 2012).
- **Boosting:** The same learning algorithm is run in the dataset under different iterations. During the first iteration, all elements of the dataset have the same weight and the algorithm tackles these elements, calculating the model error and finding the sources of error. A higher weight is allocated to these error data and the algorithm is applied once again. These iterations are repeated and the final prediction is obtained through a pondered voting process with the model weights (Ruiz, 2018).
- **Random Forest:** This method generates data subsets randomly. Each dataset defines the main attribute or characteristic of the data and divides classes based on the value of attributes. The algorithm is applied to each subset, generating a forest of trees in which the final prediction is the one that received the most votes (Bourel, 2012).
- **Stacking:** It has the particularity of combining predictions derived from different learning algorithms. The first measure consists on choosing a number M of learning algorithms that are applied to the data, then the original dataset is divided into different subsets and a learning algorithm is applied to each subset. The results of the original datasets are analyzed by a new learning algorithm and the final prediction is obtained (Jairo et al., 2017).

3. METHODOLOGY

The important questions that need to be solved in this bibliographical review are: How can we use ensemble algorithms to develop academic performance prediction models in higher education? Based on the review, which factors affect the early determination of academic performance in higher education?

3.3 Information Search

The emphasis of this work is the search of bibliographical material of the last decade using Science Direct, Springer Link, Scopus, IEEE Xplorer and Google Scholar, with keywords such as academic performance, academic performance prediction, Educational Data Mining and Ensemble Methods.

3.4 Information Organization

The reviewed information contained close to 150 articles and was filtered focusing on the ensemble methods and the factors

considered in academic performance algorithms. The first analysis delivered a total of 65 articles to be analyzed and classified into academic, sociodemographic, psychosocial and online learning.

4. ENSEMBLE METHODS FOR ACADEMIC PERFORMANCE PREDICTION

Table 1 shows a general summary of the research found as well as the algorithms used such as supervised learning and ensemble methods. Different work discussing classical algorithms were included since they often involve ensemble method implementations. The algorithms mentioned in the research articles were Neural Networks (NN); Random Forest (RF); Support Vector Machines (SVM); Naive Bayes (NB); Multilayer Perceptron (MLP); Decision Tree (DT); Online Bagging (OB); Ensemble Filtering (EF); Artificial Neural Network (ANN); K Nearest Neighbors (KNN); Bagging (BA); Boosting (BO); Logistic Regression (LOR); Linear Regression (LIR); Regression Gaussiana (RGA); Random Tree (RT); Bayesian Network (BN); Genetic Algorithm (GA); Logistic Model Trees (LMT); Decision Stump (DS); Multiple Regression (MR); Proposed model based on Logistic Regression (LOGIT_Act); Ripper (RIP); Winnow (WI); Sequential Minimal Optimization (SMO); Decision Forest (DF); Stochastic Gradient Descent (SGD); Gradient Boosting (GBT); Gaussian Bayes (GB); Stacking (ST); Long Short Term Memory (LSTM).

Table 1: Research articles found between 2010-2020 (Authors)

Year	Authors	Ensemble methods	Sample size	Algorithm with the best results	Software used
2010	(Kotsiantis et al., 2010)	WI, NB, RBF, RF, C4.5, KNN, RIP	1347	ONLINE ENSEMBLE 81,81%	-
	(Delen, 2010)	ANN, DT, SVM, LOR, BA, BO	16066	SVM 87,23%	-
2011	(Zimmermann et al., 2011)	RF	176	RF	-
	(Koutina & Kermanidis, 2011)	J48, KNN, NB, RF, SMO, Jrip	117	NB 100%	WEKA
2012	(Pardos et al., 2012)	NN, RF, ADABO, LOGIT, SWR, LR,	5422	NN 76,93%	-
2013	(Harb et al., 2013)	ADA+AG, J48, VFI, KM, ADABO	13939	ada+ag 82,07%	WEKA, Matlab
	(Romero et al., 2013)	LogR, NN, RF, NB, BN, SMO	438	MaxLogitBoost 64,85%	WEKA
2014	(Pandey & Taruna, 2014)	RTF, ADADS, BEGGREP, BEGGJ48, RF	1000	Rotation Forest 75,95%	WEKA
	(Oskouei & Askari, 2014)	NB, C4.5, RF, NN, BA, BO, ADAB, LIR, SMO	300	C 4.5 (96%)	WEKA, Clementine, Rapidminer
2015	(Strech et al., 2015)	KNN, RF, AB, SVM, NB, CT	5779	RF, AB	-
	(Bravo Sanzana et al., 2015)	RF	221189	RF	-
	(Mgala & Mbogho, 2015)	LL, MLP, LR, J48, RF, BN, NB, SMO	2426	RL 83,63%	WEKA
	(Ghatasheh, 2015)	J48, RF, MLP, SL, BN, NB, SVM.	403	SVM, SL, MLP 98.6%	WEKA
	(Goga et al., 2015)	RIP, REPTREE, DS, RF	7500	RF 99,908%	WEKA
	(Lakkaraju et al., 2015)	RD, AB, LR, SVM, DT	200000	RF	-

2016	(Stapel et al., 2016)	AB, LOR, RF, KNN, NB, SGD	566	AB 71,5%	-
	(Amrieh et al., 2016)	DT, ANN, NB, BA, BO, RF	500	ANN+BO 79,1%	WEKA
	(Kaur & Kaur, 2016)	J48, NB, REPTree, CART, ADAB	1000	Adaboost + Cart 99,60%	WEKA
	(Marbouti et al., 2016)	LOR, KNN, MLP, DT, SVM, NBC, Author Ensemble	1650	KNN 94,9%	-
	(Satyanarayana & Ravichandran, 2016)	DT, NB, RF, VO, BA, EF	1044	EF 95%	-
	(Nespereira et al., 2016)	SVM, RF	323	RF	-
2017	(Han et al., 2017)	ADAB, DT, NN, RF, SVM	119	ADAB 91,67%	-
	(Hoffait & Schyns, 2017)	NN, RF, LOGIT	6845	RF 90.2%	R
	(Abubakar et al., 2017)	NB, KNN, DT, RF	26	NB 92,3%	MATLAB
	(QIN et al., 2017)	BO, DT, RF, VO	379	RF 83,82%	R
	(Radwan & Cataltepe, 2017)	LR, RF, SVM, ADAB	649	RF	-
	(Rovira et al., 2017)	LOR, GB, SVM, RF, ADAB	4434	RF, ADAB 90%	Python
	(Asif et al., 2017)	DT, NB, NN, KNN, RF	210	NB 83,65%	RapidMiner
2018	(Al-Obeidat et al., 2018)	RF, NB, BA, LOR, Proposed classifier	-	Proposed classifier 82,2%	WEKA
	(Kumari et al., 2018)	ID3, NB, KNN, SVM, VO	500	VO 89%	WEKA
	(Hasibur Rahman & Rabiul Islam, 2018)	NB, KNN, DT, ANN, BO, B, RF	480	Adaboosting ANN 78,6%	-
	(Adejo & Connolly, 2018)	ST, DT, NN, SVM	141	ST 79,62%	SPSS, Rapid miner
	(Miguéis et al., 2018)	DT, SVM, NB, BA, BO, RF	2459	RF 96,1%	-
	(Hussain et al., 2018)	J49, RF, PART, NB	300	RF 99%	WEKA
	(Santoso, 2018)	C4.5, RF, NB, SVM	3136	RF 90.11%	-
	(Kaunang & Rotikan, 2018)	DT, RF,	249	DT 66,85%	WEKA
	(N. S. Ahmed & Sadiq, 2018)	RF	73	RF 83,56%	WEKA, C#
2019	(S.-S. M. Ajibade et al., 2019)	DT, KNN, SVM, BA, BO, RF	500	DT+BO; RF 89%	-
	(Sekeroglu et al., 2019))	SVR, LSTM, BP, GBT	-	BP 87,78%	-
	(Yekun & Teklay, 2019)	SVM, RF, KNN y MLP	893	RF 40,2%	Python
	(Adekitan & Salau, 2019)	PNN, RF, DT, NB, TE, LOR	1841	LOR 89,15%	Matlab, KNIME
	(Aman et al., 2019)	LMT, RF, J48	1021	RF 83.3%	WEKA
	(Adekitan & Noma-Osaghae, 2019)	RF, TE, DT, NB, LOR, BP, MLP	-	LOR 52%	KNIME, ORANGE
	(Tasnim et al., 2019)	J48, BO, RT, BO, RF, LMT, ADABO	395	RF 93.9%	WEKA

	(Nieto et al., 2019)	DT, RF, LOR	6100	RF 84,11%	PYTHON
	(Suguna et al., 2019)	LOR, KNN, SVM, NB, DT, RF	-	RF 91%	PYTHON
	(Fernandes et al., 2019)	GBM,	238575	GBM	H2O
	(Hassan et al., 2019)	DT, NN, SVM, RF, BO, ADAB, BAG, VO	4413	RF + VO 75, 56%	PYTHON
	(Huang et al., 2020)	GaNB, SVC, LR, DT, RF, NN y XGBoost	-	NN 88%	-
	(Braz et al., 2019)	KNN, RF, LR, MLP, DT, SVM, VO	37	VO	-
	(Singhani et al., 2019)	LOR, SVM, DT, RF	-	RF 57%	PYTHON
	(Terawaki et al., 2019)	DL, RF, SVM, NB	480	RF 77%	PYTHON, KERAS
	(Ljubobratović & Matetić, 2019)	RF	408	RF 96,3%	R
2020	(Farissi et al., 2020)	DT, NN, RF, VO, BA, BO, GA	480	GA+RF, 81,64%	-
	(Ashraf et al., 2020)	J48, RT, NB, KNN, BO	-	J48 + BO: 96,5%	-
	(M. N. Injadat et al., 2020)	RF, SVM - RBF, k-NN, NB, LR y NN	486	BA 93,1%	MATLAB, R
	(M. Injadat et al., 2020)	KNN, RF, SVM, LOR, MLP, NB	52	NB+LOR 96.6%	R
	(Rincon-Flores et al., 2020)	KNN, RF	134	RF	-
	(Hassan et al., 2020)	RF, BA, ADAB, GMB, XGBOOST	4413	ADAB 71.8%	PYTHON, JUPYTER
	(Patacsil, 2020)	DT, FT, J48, VO, BA, ADAB	2401	BA+J48: 70,47%	RapidMiner
	(Jayaprakash et al., 2020)	NB, BO, RF,	887	RF 93%	WEKA
	(S. S. M. Ajibade et al., 2020)	NB, DT, KNN, ADAB, BA, BO	500	ADAB + NB: 86%	MATLAB
	(Ganatra & Nilkant, 2020)	CT, BA, RF	215	RF 92,59%	R
	(M. R. Ahmed et al., 2020)	KNN, NB, BA, RF, J48	800	KNN 91,37%	WEKA, MATLAB
	(Wu & Wu, 2020)	RF, SVM, Ridge Linear Regression	596	SVM	WEKA

Based on the literature review, the most commonly used algorithms in decreasing order are RF, NB, DT, SVM and NN. Although the ensemble method random forest (RF) was the most used in literature, ML techniques are still used. However, there has been a progressive increase in the use of ensemble methods, especially over the last five years, to explore the possible causes of different academic issues such as performance. Table 2 presents the usage frequency of the machine learning algorithms from 2016 to 2020. It is important to highlight that, in 45 of the articles found, the authors stated the academic performance predictions had higher accuracy with an ensemble algorithm. The Random Forest method is well known and applied in the field, followed by Adaboost, Bagging, Boosting, voting and XGBoosting.

Table 2: Usage of ensemble algorithms over the last five years. Source: Authors

	AdaBoost	Bagging	Boosting	Random Forest	Voting	XGBoost
2020	3	6	4	9	2	1
2019	2	2	4	14	2	1
2018		2	2	7	1	
2017	3		1	7	1	
2016	2	2	1	4	1	

5. FACTORS THAT AFFECT ACADEMIC PERFORMANCE

According to the literature review, authors classify the studied factors in one of the following categories:

ACADEMIC: These are the most studied factors since they come from the information delivered by higher education institutions to assess the learning level of students, and allows predictions on the academic performance, with proper accuracy (Eckert & Suénaga, 2015). This category of factors includes information on average grades in previous institutions, grades in tests and subjects, admission score and GPA.

DEMOGRAPHICAL: In this category, social aspects and general characteristics of students are considered (Contreras et al., 2020). The information classified as demographical encompasses age, gender, nationality, language, civil status, number of family members, education of the parents, guardian/responsible adult, place of residence, etc. Usually, authors fill out this information by handing polls to students and relatives.

E-LEARNING: Technological progress and the possibility to access information online has led institutions to implement online activities and tools that can complement the classes received. In this category, authors considered attributes related to the behavior of students that use virtual education platforms, such as the number of logins, number of homework sent, number of forums visited, comments viewed, online hours, total number of documentation files reviewed, etc.

PSYCHOSOCIAL FACTORS: Psychosocial factors comprise motivation, time dedicated to academic activities, student behavior, participation in class, communicational skills and, in general, the behavior type of students when they interact with their environment. Many authors analyze how motivation, interest and personal goals reflects on their academic performance.

Figure 2: Number of articles studied per factor. Source: Authors

6. DISCUSSION OF RESULTS

In 52 of the analyzed studies, the factors linked to the academic setting are the most reviewed attributes to predict the variables that affect the academic performance of students. GPA, scores obtained throughout the academic semester and scores in individual tests are amongst the most relevant attributes partly because this information is made accessible by education entities. Furthermore, the student attributes related with demographic characteristics are also important in academic performance studies.

At least 37 out of the 65 research projects analyzed consider characteristics such as gender, age, residence, family size, education of parents, which are widely used since they encompass various influence variables. This information is usually collected through polls answered by students.

Regarding psychosocial behavior factors such as motivation and personal interest were considered in 25 of the reviewed articles while factors linked to the learning process appeared in 14 referenced articles. The advantages of online learning have nourished the search of variables that can measure the performance of students. Other aspects include number of logins, number of homework assignments sent, number of tests, scoring activities, time spent on academic platforms (Figure 2).

7. CONCLUSIONS

The academic performance is an essential element in higher education quality, since this indicator leads to an approximation of reality in the education sector and affects other variables such as dropout and university abandonment.

The literature review shows that the most commonly used algorithms are RF, NB, DT, SVM and NN. These have been increasingly adopted over the past five years, with the exploration of ensemble methods to explain the causes of many academic problems such as low performance in students.

Many software tools are being used to run ensemble and ML algorithms such as Weka, Matlab, Python and R. These involve variables that can be mostly grouped within the sociodemographic and academic aspects, followed by psychosocial and e-learning factors.

REFERENCES

1. Abubakar, Y., Bahiah, N., & Ahmad, H. (2017). Prediction of Students' Performance in E-Learning Environment Using Random Forest. In *International Journal of Innovative Computing* (Vol. 7, Issue 2).
2. Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. *Journal of Applied Research in Higher Education*, 10(1), 61–75.
3. Adekitan, A. I., & Noma-Osaghae, E. (2019). Data mining approach to predicting the performance of first year student in a university using the admission requirements. *Education and Information Technologies*, 24(2), 1527–1543.
4. Adekitan, A. I., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon*, 5(2), e01250.
5. Ahmed, M. R., Tahid, S. T. I., Mitu, N. A., Kundu, P., & Yeasmin, S. (2020). A Comprehensive Analysis on Undergraduate Student Academic Performance using Feature Selection Techniques on Classification Algorithms. 1–6.
6. Ahmed, N. S., & Sadiq, M. H. (2018). Clarify of the random forest algorithm in an educational field. 2018 *International Conference on Advanced Science and Engineering (ICOASE)*, 179–184.
7. Ajibade, S.-S. M., Ahmad, N. B. B., & Shamsuddin, S. M. (2019). Educational Data Mining: Enhancement of Student Performance model using Ensemble Methods. *IOP Conference Series: Materials Science and Engineering*, 551(1), 12061.
8. Ajibade, S. S. M., Ahmad, N. B., & Shamsuddin, S. M. (2020). A data mining approach to predict academic performance of students using ensemble techniques. *Advances in Intelligent Systems and Computing*, 940, 749–760.
9. Al-Obeidat, F., Tubaishat, A., Dillon, A., & Shah, B. (2018). Analyzing students' performance using multi-criteria classification. *Cluster Computing*, 21(1), 623–632.
10. Aman, F., Rauf, A., Ali, R., Iqbal, F., & Khattak, A. M. (2019). A Predictive Model for Predicting Students Academic Performance. 10th *International Conference on Information, Intelligence, Systems and Applications, IISA 2019*, 1–4.
11. Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining educational data to predict student's academic performance using ensemble methods. *International Journal of Database Theory and Application*, 9(8), 119–136.
12. Ashraf, M., Zaman, M., & Ahmed, M. (2020). An Intelligent Prediction System for Educational Data Mining Based on Ensemble and Filtering approaches. *Procedia Computer Science*, 167, 1471–1483.
13. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education*, 113, 177–194. <https://doi.org/10.1016/j.compedu.2017.05.007>
14. Bourel, M. (2012). Model aggregation methods and applications. 10, 19–32.
15. Bravo Sanzana, M., Salvo Garrido, S., & Muñoz Poblete, C. (2015). Profiles of Chilean students according to academic performance in mathematics: An exploratory study using classification trees and random forests. *Studies in Educational Evaluation*, 44, 50–59.
16. Braz, F., Campos, F., Stroele, V., & Dantas, M. (2019). An Early Warning Model for School Dropout: a Case Study in E-learning Class. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática Na Educação-SBIE)*, 30(1), 1441.
17. Bunce, D. M., & Hutchinson, K. D. (1993). The use of the GALT (Group Assessment of Logical Thinking) as a predictor of academic success in college chemistry. ACS Publications.
18. Caballero, C. C., & Palacio, J. (2007). Relación del burnout y el rendimiento académico con la satisfacción frente a los estudios en estudiantes universitarios. *Avances En Psicología Latinoamericana*, 25(2), 98–111.

19. Contreras, L., Fuentes, H., & Rodriguez, J. (2020). Application of automatic learning as a prediction strategy for academic dropout in universities. *Sylwan Journal*, 164(6).
20. Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, 49(4), 498–506.
21. Eckert, K. B., & Suénaga, R. (2015). Análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en minería de datos. *Formacion Universitaria*, 8(5), 3–12.
22. Farissi, A., Mohamed Dahlan, H., & Samsuryadi. (2020). Genetic Algorithm Based Feature Selection With Ensemble Methods For Student Academic Performance Prediction. *Journal of Physics: Conference Series*, 1500, 12110.
23. Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Erven, G. Van. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, 94, 335–343.
24. Ganatra, D., & Nilkant, D. (2020). Ensemble Methods to Improve Accuracy of a Classifier. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(3), 3434–3439.
25. García-Tinisaray, D., Manuel, J., & Pichardo, M. (2019). Learning Analytics as an analysis factor of university academic performance. In E. S. P. del L. (ESPOL) (Ed.), *CEUR Workshop Proceedings*, (pp. 42–50).
26. García, Y., López de Castro, D., & Rivero, O. (2014). Estudiantes universitarios con bajo rendimiento académico, ¿qué hacer? *Revista Edumecentro*, 6(2), 272–278.
27. Ghatasheh, N. (2015). Knowledge level assessment in e-learning systems using machine learning and user activity analysis. *Article in International Journal of Advanced Computer Science and Applications*, 6(4), 107–113.
28. Goga, M., Kuyoro, S., & Goga, N. (2015). A Recommender for Improving the Student Academic Performance. *Procedia - Social and Behavioral Sciences*, 180, 1481–1488.
29. Han, M., Tong, M., Chen, M., Liu, J., & Liu, C. (2017). Application of Ensemble Algorithm in Students' Performance Prediction. *Proceedings - 2017 6th IIAI International Congress on Advanced Applied Informatics, IIAI-AAI 2017*, 735–740.
30. Harb, H. M., Emara, A. H., Elden, A., Moustafa, M. A., & Emara, A. (2013). Adaboost Ensemble with Simple Genetic Algorithm for Student Prediction Model. *International Journal of Computer Science & Information Technology (IJCSIT)*, 5(2).
31. Hasibur Rahman, M., & Rabiul Islam, M. (2018). Predict Student's Academic Performance and Evaluate the Impact of Different Attributes on the Performance Using Data Mining Techniques. *2nd International Conference on Electrical and Electronic Engineering, ICEEE 2017*, 1–4.
32. Hassan, H., Ahmad, N. B., & Anuar, S. (2020). Improved students' performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining. *Journal of Physics*, 1529, 52041.
33. Hassan, H., Anuar, S., & Ahmad, N. B. (2019). Students' performance prediction model using meta-classifier approach. *Communications in Computer and Information Science*, 1000, 221–231. https://doi.org/10.1007/978-3-030-20257-6_19
34. Hoffait, A. S., & Schyns, M. (2017). Early detection of university students with potential difficulties. *Decision Support Systems*, 101, 1–11.
35. Huang, A. Y. Q., Lu, O. H. T., Huang, J. C. H., Yin, C. J., & Yang, S. J. H. (2020). Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs. *Interactive Learning Environments*, 28(2), 206–230.

36. Hussain, S., Dahan, N. A., Najoua, R., Mutaheer Ba-Alwib, F., & Ribata, N. (2018). Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, 9(2), 447–459.
37. Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2020). Multi-split optimized bagging ensemble model selection for multi-class educational data mining. *Applied Intelligence*. <https://doi.org/10.1007/s10489-020-01776-3>
38. Injadat, M. N., Moubayed, A., Nassif, A. B., & Shami, A. (2020). Systematic ensemble model selection approach for educational data mining. *Knowledge-Based Systems*, 200, 105992.
39. Jairo, J., Yepes, C., Leonardo, D., & Castro, C. (2017). Modelos Apilados y factores que pueden afectar la eficiencia Stacked Models and factors that can affect efficiency.
40. Jayaprakash, S., Krishnan, S., & Jaiganesh, J. (2020). Predicting Students Academic Performance using an Improved Random Forest Classifier. *2020 International Conference on Emerging Smart Computing and Informatics, ESCI 2020*, 238–243.
41. Kaunang, F. J., & Rotikan, R. (2018). Students' academic performance prediction using data mining. *Proceedings of the 3rd International Conference on Informatics and Computing, ICIC 2018*, 1–5.
42. Kaur, K., & Kaur, K. (2016). Analyzing the effect of difficulty level of a course on students performance prediction using data mining. *Proceedings on 2015 1st International Conference on Next Generation Computing Technologies, NGCT 2015*, 756–761.
43. Kotsiantis, S., Patriarcheas, K., & Xenos, M. (2010). A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. *Knowledge-Based Systems*, 23(6), 529–535.
44. Koutina, M., & Kermanidis, K. L. (2011). Predicting postgraduate students' performance using machine learning techniques. *IFIP Advances in Information and Communication Technology*, 364 AICT(PART 2), 159–168.
45. Kumari, P., Jain, P. K., & Pamula, R. (2018). An efficient use of ensemble methods to predict students academic performance. *Proceedings of the 4th IEEE International Conference on Recent Advances in Information Technology, RAIT 2018*, 1–6.
46. Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1909–1918.
47. Ljubobratović, D., & Matetić, M. (2019). Using LMS Activity Logs to Predict Student Failure with Random Forest Algorithm. *The future of information sciences*, 113.
48. Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers and Education*, 103, 1–15. <https://doi.org/10.1016/j.compedu.2016.09.005>
49. Mgala, M., & Mbogho, A. (2015). Data-driven intervention-level prediction modeling for academic performance. *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development - ICTD '15*, 15, 1–8.
50. Miguéis, V. L., Freitas, A., Garcia, P. J. V., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, 115, 36–51.
51. Moreno, H., & Chauta, L. (2012). Funcionalidad familiar, conductas externalizadas y rendimiento académico en un grupo de adolescentes de la ciudad de Bogotá. *Rev Psychol Av Discip*, 6(1), 155–166.
52. Nespereira, C. G., Elhariri, E., El-Bendary, N., Vilas, A. F., & Redondo, R. P. D. (2016). Machine learning based classification approach for predicting students performance in blended learning. *The 1st International Conference on Advanced Intelligent*

- System and Informatics (AISI2015)*, November 28-30, 2015, Beni Suef, Egypt, 47–56.
53. Nieto, Y., Garcia, V., Montenegro, C., Gonzalez, C., & Gonzalez, R. (2019). *Usage of Machine Learning for Strategic Decision Making at Higher Educational Institutions*. *IEEE Access*, 7, 75007–75017.
 54. Oskouei, R. J., & Askari, M. (2014). *Predicting academic performance with applying data mining techniques (generalizing the results of two different case studies)*. *Computer Engineering and Applications Journal*, 3(2), 79–88.
 55. Pandey, M., & Taruna, S. (2014). *A Comparative Study of Ensemble Methods for Students' Performance Modeling Aspect level sentiment analysis*. *Article in International Journal of Computer Applications*, 103(8), 975–8887.
 56. Pardos, Z. A., Gowda, S. M., Baker, R. S. J. d., & Heffernan, N. T. (2012). *The Sum is Greater than the Parts: Ensembling Models of Student Knowledge in Educational Software*. *ACM SIGKDD Explorations Newsletter*, 13(2), 37–44.
 57. Patacsil, F. F. (2020). *Predicting University Students' Academic Success Using Different Tree Classifiers And Ensemble Approaches To Suggest Suitable Program*. *International journal of scientific & technology research*, 9(02), 6001–6009.
 58. Qin, f., zhu, l., cheng, z., & zhang, q. (2017). *Research on Student Performance Evaluation Based on Random Forest*. *DEStech Transactions on Engineering and Technology Research*, eeta, 387–393.
 59. Radwan, A. M., & Cataltepe, Z. (2017). *Improving Performance Prediction on Education Data with Noise and Class Imbalance*. *Intelligent Automation & Soft Computing*, 1–8.
 60. Rincon-Flores, E. G., Lopez-Camacho, E., Mena, J., & Lopez, O. O. (2020). *Predicting academic performance with Artificial Intelligence (AI), a new tool for teachers and students*. *IEEE Global Engineering Education Conference, EDUCON, 2020-April*, 1049–1054.
 61. Rojas, M. E. (2008). *Introduction to machine learning isabelle guyon 2008-02-31 Notas tomadas por: María Eugenia Rojas Qué es Machine Learning?*
 62. Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). *Web usage mining for predicting final marks of students that use Moodle courses*. *Computer Applications in Engineering Education*, 21(1), 135–146.
 63. Rovira, S., Puertas, E., & Igual, L. (2017). *Data-driven system to predict academic grades and dropout*. *PLoS One*, 12(2),
 64. Ruiz, J. Z. (2018). *Comparativa y análisis de algoritmos de aprendizaje automático para la predicción del tipo predominante de cubierta arbórea*. *Universidad complutense de madrid*.
 65. Russo, C., Ramón, H. D., Alonso, N., Cicerchia, L. B., Esnaola, L., & Tessore, J. P. (2016). *Tratamiento masivo de datos utilizando técnicas de Machine Learning*. *XVIII Workshop de Investigadores En Ciencias de La Computación (Entre Ríos, Argentina)*.
 66. Salcedo, A. (2010). *Desertion in Colombian Universities*. *Revista Academia y Virtualidad*, 3(1), 50–60.
 67. Santoso, L. W. (2018). *Early Warning System for Academic using Data Mining*. *Proceedings - 2018 4th International Conference on Advances in Computing, Communication and Automation, ICACCA 2018*, 1–4.
 68. Satyanarayana, A., & Ravichandran, G. (2016). *Mining Student data by Ensemble Classification and Clustering for Profiling and Prediction of Student Academic Performance*. *American Society for Engineering Education*.
 69. Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019). *Student performance prediction and classification using machine learning algorithms*. *ACM International Conference Proceeding Series, Part F148151*, 7–11.
 70. Singhani, S., Desai, S., Bailurkar, R., & Mantri, R. (2019). *Student Academic Performance Prediction using Machine Learning*. *International journal of information and computing science*, 6(4), 234–241.

71. Stapel, M., Zheng, Z., & Pinkwart, N. (2016). *An Ensemble Method to Predict Student Performance in an Online Math Learning Environment*. International Educational Data Mining Society.
72. Strecht, P., Cruz, L., Soares, C., Mendes-Moreira, J., & Abreu, R. (2015). *A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance*. *Proceedings of the 8th International Conference on Educational Data Mining*, 392–395.
73. Suguna, R., Shyamala Devi, M., Bagate, R. A., & Joshi, A. S. (2019). *Assessment of feature selection for student academic performance through machine learning classification*. *Journal of Statistics and Management Systems*, 22(4), 729–739.
74. Tasnim, N., Paul, M. K., & Sarowar Sattar, A. H. M. (2019). *Performance Analysis of Different Decision Tree Based Methods for Identifying Drop out Students*. *1st International Conference on Advances in Science, Engineering and Robotics Technology 2019, ICASERT 2019, 2019(Icasert)*, 1–6.
75. Terawaki, Y., Unoki, T., Kato, T., & Kodama, Y. (2019). *Performance evaluation for four types of machine learning algorithms using educational open data*. *Smart Innovation, Systems and Technologies*, 144, 281–289.
76. Tian, H., & Sun, Z. (2018). *Academic Achievement Assessment: Principles and Methodology*. In *Academic Achievement Assessment*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-56198-0_1
77. Velázquez, L. E. T., & Soriano, N. Y. R. (2006). *Rendimiento académico y contexto familiar en estudiantes universitarios*. *Enseñanza e Investigación En Psicología*, 11(2), 255–270.
78. Wu, X., & Wu, J. (2020). *Criteria evaluation and selection in non-native language MBA students admission based on machine learning methods*. *Journal of Ambient Intelligence and Humanized Computing*, 11(9), 3521–3533.
79. Yekun, E. A., & Teklay, A. (2019). *Student Performance Prediction with Optimum Multilabel Ensemble Model*. *ArXiv Preprint ArXiv:1909.07444*.
80. York, T. T., Gibson, C., & Rankin, S. (2015). *Defining and measuring academic success*. *Practical Assessment, Research, and Evaluation*, 20(1), 5.
81. Zimmermann, J., Zurich, E., Brodersen, K. H., Buhmann, J. M., Zimmermann, J., Brodersen, K. H., Pellet, J.-P., August, E., & Buhmann, J. M. (2011). *Predicting Graduate-level Performance from Undergraduate Achievements Whole-brain dynamic causal modeling of fMRI data View project Bayesian inference in sparse settings View project Predicting graduate-level performance from undergraduate achievements*.