



Article

A Machine Learning Approach to Predicting Academic Performance in Pennsylvania's Schools

Shan Chen ¹ and Yuanzhao Ding ^{2,*}

¹ Department of Applied Social Sciences, The Hong Kong Polytechnic University, 11 Yuk Choi Rd., Hung Hom, Hong Kong, China

² School of Geography and the Environment, University of Oxford, South Parks Road, Oxford OX1 3QY, UK

* Correspondence: armstrongding85@gmail.com

Abstract: Academic performance prediction is an indispensable task for policymakers. Academic performance is frequently examined using classical statistical software, which can be used to detect logical connections between socioeconomic status and academic performance. These connections, whose accuracy depends on researchers' experience, determine prediction accuracy. To eliminate the effects of logical relationships on such accuracy, this research used 'black box' machine learning models extended with education and socioeconomic data on Pennsylvania to predict academic performance in the state. The decision tree, random forest, logistic regression, support vector machine, and neural network achieved testing accuracies of 48%, 54%, 50%, 51%, and 60%, respectively. The neural network model can be used by policymakers to forecast academic performance, which in turn can aid in the formulation of various policies, such as those regarding funding and teacher selection. Finally, this study demonstrated the feasibility of machine learning as an auxiliary educational decision-making tool for use in the future.

Keywords: machine learning; neural network; socioeconomic status; population; crime rate; academic performance



Citation: Chen, Shan, and Yuanzhao Ding. 2023. A Machine Learning Approach to Predicting Academic Performance in Pennsylvania's Schools. *Social Sciences* 12: 118. <https://doi.org/10.3390/socsci12030118>

Academic Editor: Neil Gilbert

Received: 21 December 2022

Revised: 14 February 2023

Accepted: 15 February 2023

Published: 24 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

For a considerable duration, educational systems have widely utilized standardized examinations as a large-scale means of effectively sorting students. When it comes to evaluation efficiency, standardized test scores are overwhelmingly superior in identifying talent over other qualities that schools ought to place greater emphasis on, such as moral character, life adaptability, non-cognitive skills, and social responsibility (Ebel and Frisbie 1972). These preferences are rooted in the strengths of standardized tests, which are a product of historical and social conventions. There are considerable and obvious advantages to employing paper-and-pencil examinations that feature a series of archetypal questions, including practicality, reliability, good content validity, convenience, accessibility, and openness.

Despite the usefulness of traditional tests in assessing students' knowledge and skills, there are several other factors that can impact academic performance, often overlooked. One significant factor identified in predictive studies is socioeconomic status (SES), which plays a vital role in widening the academic performance gap between students in rural and urban institutions (Ramos et al. 2012). In some European countries, high SES often correlates with above-average exam scores, highlighting the significant impact of SES on educational performance (Jana et al. 2006; Willms et al. 2006). Conversely, in eastern Europe, low SES and students from rural schools can negatively affect academic performance (Kryst et al. 2015). While some studies found few differences between rural and urban schools (Miller et al. 2019) or no significant distinctions among students from different school settings (Fan and Chen 1998). Many researchers and educators continue to explore the effects of SES

on academic performance using correlation and regression analysis. As such, this study aims to employ machine learning (ML) models, a novel approach in predictive studies, to investigate the impact of SES on student academic performance. Numerous studies have demonstrated the considerable accuracy of ML prediction compared to a classic statistical method such as correlation and linear regression (Table 1) (Chang et al. 2020; Paulick et al. 2013). As an artificial intelligence approach, ML has had a far-reaching influence on handling the vast amounts of facts and numerical data generated by computers through simulations of the human brain. For instance, an ML algorithm is superior in analyzing considerable internet data than regular models, since it enables relatively rapid prediction with high accuracy and large datasets (Fedushko and Ustyianovych 2019; Shakhovska et al. 2017; Zhou et al. 2017). Applying ML algorithms also enables researchers and teachers to recognize the key factors that strongly influence student performance and find more effective ways to improve teaching quality (Buenaño-Fernández et al. 2019; Hussain et al. 2019; Kemper et al. 2020). The problem is that previous studies were a small-scale, incomprehensive and restricted data pool to address certain groups under limited conditions. This scope cannot ensure overall effective outcomes of ML prediction, and a large representative sample has yet to be used to further verify the precision of ML results.

Table 1. Comparison of a classical statistical method (e.g., correlation and linear regression) and ML.

	Classical Statistical Method	ML Method	Reference
Rationale	Necessary for understanding the relationship between academic performance and relevant factors (e.g., crime rate and population density)	Prediction of academic performance by ML algorithms	(Bujang et al. 2021; Chang et al. 2020; Lykourantzou et al. 2009; Mduma et al. 2019; Papernot et al. 2017; Paulick et al. 2013; Şara et al. 2015)
Methods	The use of programs such as Mplus to identify relationships between academic performance and relevant factors; calculation based on the relationships	Prediction via ‘black box’ models without consideration for relationships	(Bujang et al. 2021; Chang et al. 2020; Lykourantzou et al. 2009; Mduma et al. 2019; Papernot et al. 2017; Paulick et al. 2013; Şara et al. 2015)
Accuracy	Existing relationships and assumptions	Quality and quantity of data	(Al-Jarrah et al. 2015; Ciolacu et al. 2017; Sekeroglu et al. 2019)
Advantages	Matured methods with clear processes	Rapid and convenient prediction for reasonable results	(Ciolacu et al. 2017; Sekeroglu et al. 2019)
Limitations	Sample selection bias	The ‘black swan’ effect	(Batrouni et al. 2018; Lorey et al. 2011)

In ML prediction, different variables may strongly affect student performance. In this respect, Musso et al. administered a questionnaire on digital tools, health, social support, demographic items, cognitive attributes, and learning and coping strategies and used a neural network algorithm to predict student performance (Musso et al. 2020). Qazdar et al. incorporated several variables, such as gender, test score, and performance, into the forecasting of students’ test results (Qazdar et al. 2019). Yousafzai et al. use a digital management system (which reflects student information and academic progress) to their advantage in predicting test scores (Yousafzai et al. 2020). The decision tree and KNN models used by the authors achieved an accuracy of 85%. Alyahyan and Düşteğör explored the factors that contribute to successful performance in academics (e.g., sociodemographic, psychological and academic factors, and cognitive qualities) (Alyahyan and Düşteğör 2020). Boxer et al. discovered a negative relationship between crime and student performance in language and math, with impoverished students engaged in more delinquencies and criminal events (Boxer et al. 2020). Although there are many factors play a role in the prediction of student performance, sociodemographics, and crime rates have an important influence on academic performance in schools.

Above all, the present research focused on the effects of socioeconomic status (SES) and crime rates on school performance. The study chose Pennsylvania as the basis for our study and an ML model to predict academic performance in the state. Using population, crime, and school data, this study trained five ML models: a decision tree, random forest, logistic regression, support vector machine, and neural network (Figure 1). Among these models, the neural network could predict overall academic performance in schools precisely, despite the significant deviations among individual students, such as abnormal performance in examinations. This study also demonstrated the capability of the neural network to identify which factors (e.g., crime rate) are the most important in affecting academic performance. In summary, this work pointed to the feasibility of the ML model as an auxiliary tool for decision making in the future.

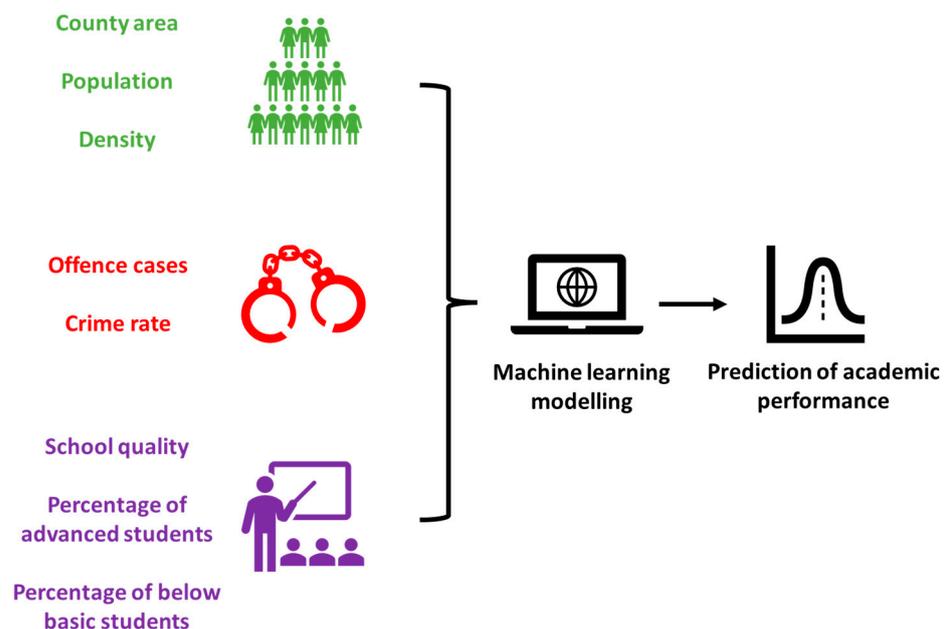


Figure 1. Framework of this study shows how ML used big data to predict academic performance.

2. Materials and Methods

2.1. Data Collection

Pennsylvania is selected as the model area because (1) Pennsylvania provides a full set of online education data; (2) Pennsylvania is a representative state that includes megacities and rural areas; and (3) Pennsylvania is a state with a wide range of educational resources. The educational data are downloaded from Pennsylvania School Performance Profile (<https://paschoolperformance.org/>, accessed on 1 December 2022), including “Grade”, “School level”, “Sample size”, “Subject”, “Percent of advanced student”, and “Percent of below-basic student”. The county data are obtained from the United States Census Bureau (<https://www.census.gov/>, accessed on 1 December 2022) and World Population Review (<https://worldpopulationreview.com/us-counties/states/pa>, accessed on 1 December 2022), such as “County area”, “County population” and “County density”. The crime data are taken from Pennsylvania uniform crime reporting system (UCR) records (<https://www.attorneygeneral.gov/>, accessed on 1 December 2022), such as “Total offence cases”, and “Crime rate”. The rural-urban definitions are referred to Center for Rural Pennsylvania (<https://www.rural.pa.gov/data/rural-urban-definitions>, accessed on 1 Dec 2022).

Based on our classifications presented in Table 2, we performed statistical analysis and ML calculations.

Table 2. Classification method for data treatment.

County Area (km ²)	Approximate Number	County Population (People)	Approximate Number	Crime Rate (per 1000 People)	Approximate Number
≤1000	500	≤50,000	25,000	(3, 6]	5
(1000, 2000]	1500	(50,000, 200,000]	100,000	(6, 10]	8
(2000, 3000]	2500	(200,000, 1,000,000]	500,000	(10, 16]	13
(3000, 4040]	3500	(1,000,000, 1,584,064]	1,500,000	(16, 30]	29
Population Density (people/km ²)	Approximate Number	Total Offenses Cases	Approximate Number	Percentage of Advanced/Below-Basic Students	Approximate Number
≤100	50	≤10,000	5,000	0%	0%
(100, 500]	300	(10,000, 50,000]	25,000	(0%, 10%]	5%
(500, 1300]	900	(50,000, 200,000]	100,000	(10%, 20%]	15%
(1300, 4564]	3000	(200,000, 859,411]	500,000	(20%, 40%]	30%
				(40%, 60%]	50%
				(60%, 100%]	80%
Subject	Assigned Number	School Level	Assigned Number	Rural/Urban	Assigned Number
English language	1	Historically under performance	1	Rural	1
Math	2	All group	2	Urban	2
Science	3				

2.2. ML Models

More than 33,000 educational records were input into the ML models. Unless otherwise stated, the split ratio was 76–24% for train-test sets. After the model was trained, this study applied this model to make predictions on the data in unknown areas (see Figure 1). The authors trained the ML via Anaconda 3 and Jupyter 6.3.0 platform. The python coding library was based on scikit-learn (sklearn), keras, pandas, and matplotlib. Five ML methods were compared: decision tree (Somvanshi et al. 2016), random forest (Liu et al. 2012), logistic regression (Rymarczyk et al. 2019), support vector machine (Somvanshi et al. 2016), and neural network (Jung and Kim 2016; Qi et al. 2019). ML methods were based on previous studies with default settings in the sklearn module (Chen and Ding 2022; Pedregosa et al. 2011).

The neural network consisted of 100 hidden layers, each with 100 nodes. The maximum number of iterations was 50. The activation function was the rectified linear unit (relu). The solver for the neural network algorithm was adam optimization (“adam”). The python coding for training, testing, and prediction was also attached to the supplemental information (Pomerat et al. 2019). ML coding was attached to supplemental material Table S1. The tuning process followed the random search method (Table S2) and tuning result was attached to Table S8.

Four Pennsylvania heatmaps were drawn by RStudio: county population, total offense cases, percentage of advanced students (real situation), and percentage of advanced students (neural network prediction). The RStudio coding was based on packages of tidyverse, readr, and maps. The color bar followed terrain.colors, and heat.colors. R coding for the heatmap was attached in the Supplemental Material Tables S3–S7.

For this study, Dell Inspiron 15 TGL 3000 with Intel Core™ i7-1165G7 CPU and 16 GHz 3200 MHz memory was used unless otherwise stated. The total calculation time to run the coding was around 1–2 h for each round. The computing power of the computer significantly affected performance. High-performance computing was required when running the coding (Correa-Baena et al. 2018).

3. Results

3.1. Educational Data Analysis

The correlation heatmap was shown in Figure 2. From the correlation heatmap, we found that the population density, total offense cases, and crime rates had a strong positive correlation with each other. A higher population in the county normally implied higher density (+0.85), more total offense cases (+0.90), and a higher crime rate (+0.79).

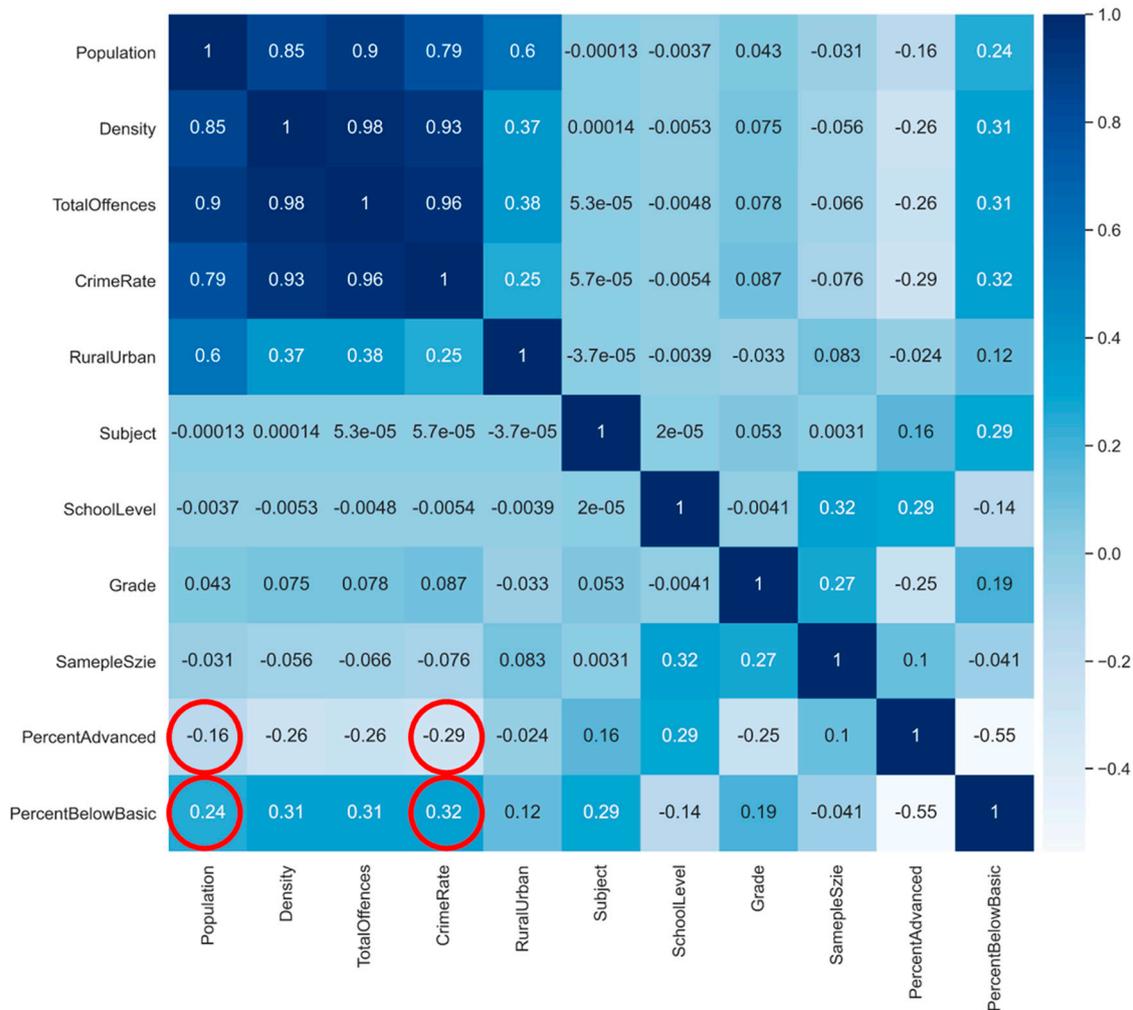


Figure 2. Heatmap analysis showing the relationship between “Area”, “Population”, “Density”, “Total offences”, “Crime rate”, “Rural or urban”, “Grade”, “School level”, “Sample size”, “Subject”, “Percentage of advanced students”, and “Percentage of below-basic students”. The darker the color, the stronger the positive correlation between the two data sets. The lighter the color, the stronger the negative correlation between the two data sets. The red circles visually represent the focal points of this paper’s target analysis, namely the relationships between “Percentage of advanced students”, “Percentage of below basic students”, “Population”, and “Crime rate”.

Based on the heatmap, the main factors affecting academic performance (using the percentage of advanced students) were population (−0.16), density (−0.26), total offense cases (−0.26), crime rate (−0.29), rural and urban (−0.024), grade (−0.25), school level (+0.29). The lower population, lower density, lower offense cases, and lower crime rate led to a higher percentage of advanced students.

On the contrary, the higher population, higher density, higher offense cases, and higher crime rate led to a higher percentage of below-basic students. The study environment had a significant impact on the overall academic performance.

Feature importance (Figure S1) results showed that the most important factor affecting the data is the sample size (+0.458), grade (+0.126), crime rate (+0.124), subject (+0.097), and population (+0.051). The sample size was the most important factor affecting the calculation since more students had a greater impact on prediction outcomes. The crime rate and population were among the top five important factors affecting the prediction result. In the following analysis, we examined the academic impact of the crime rate and population in detail.

When we evaluated how the county population affects academic performance, the authors found the large county population (1 M~1.58 M, red box and red arrow in Figure 3) led to a lower percentage of advanced students and a higher percentage of below-basic students. The larger county population suggested a lower academic performance. On the contrary, a smaller county population helped to improve students' overall academic performance.

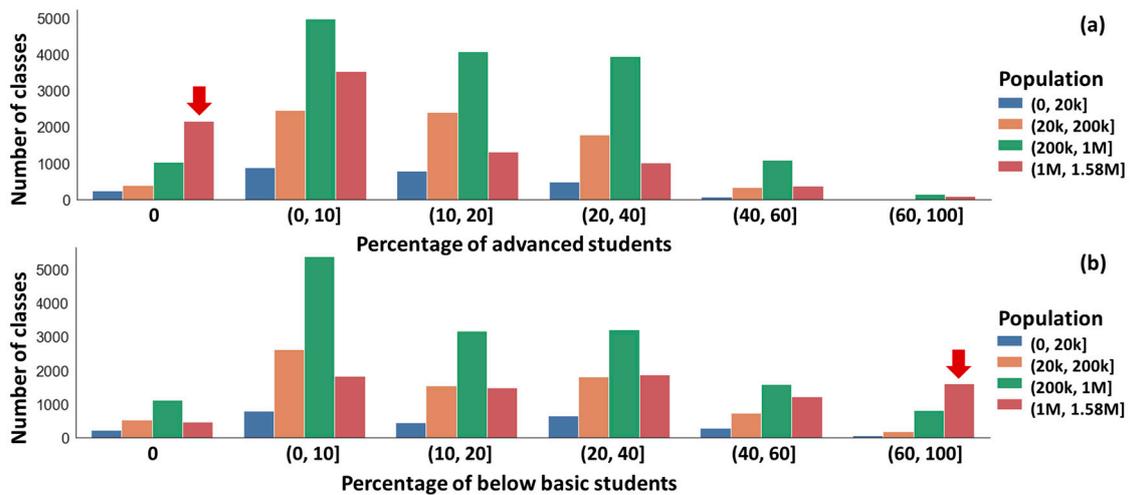


Figure 3. Analysis of the relationship between county population and academic performance; (a) percentage of advanced students; (b) percentage of below basic students. The red arrows indicate the highest population in the county, which are the groups specifically analyzed.

When the authors evaluated how a safe environment affected academic performance, the authors found the high crime rate (16–30 cases per 1000, red box and red arrow in Figure 4) led to a lower percentage of advanced students and a higher percentage of below-basic students. In summary, a higher crime rate led to lower academic performance. On the other side, safe environments led to higher academic performance.

The school level significantly affected academic performance (Figure 5a). In the historically underperforming schools, most of the classes only had around 2% of advanced students. In all other schools, most of the classes had around 4% of advanced students. Especially, all other schools had roughly twice as many excellent classes (more than 50% advanced students) as historically underperforming schools. Although historically underperforming schools also had some excellent classes and some good students, a higher school level significantly improved the overall academic performance. Rural schools had better overall academic performances compared to urban schools. Most rural schools had around 8% of the advanced students (red dashed line in Figure 5b), while most urban schools had only 2% of the advanced students (blue dashed line in Figure 5a).

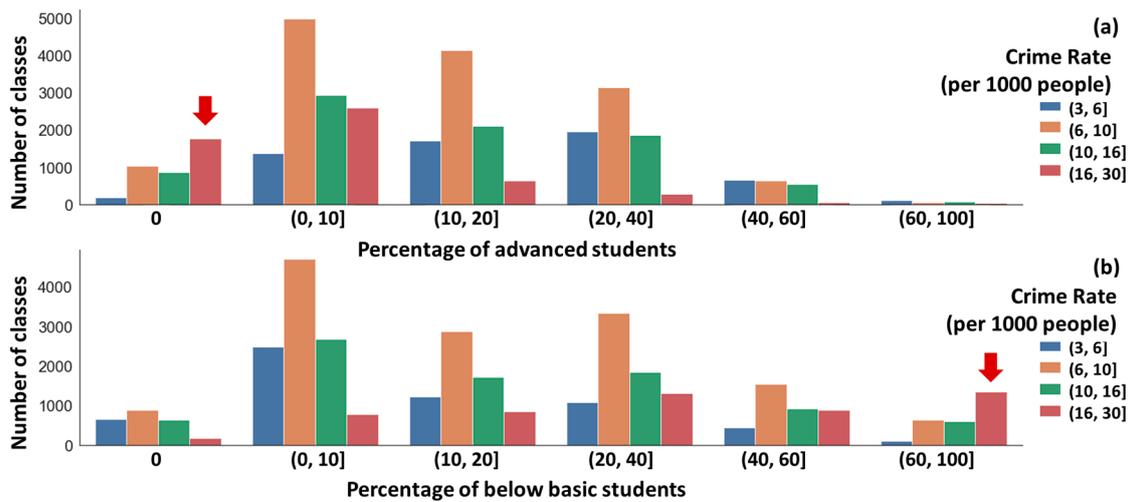


Figure 4. Analysis of the relationship between crime rate (per 1000 people) and academic performance; (a) percentage of advanced students; (b) percentage of below basic students. The red arrows indicate the highest crime rates, which are the groups specifically analyzed.

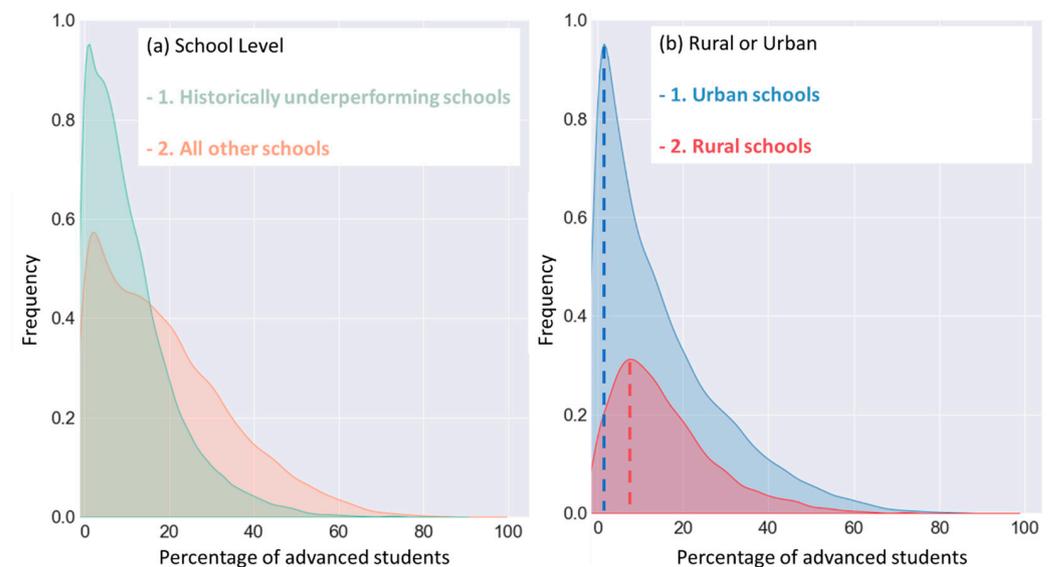


Figure 5. (a) The school level affecting academic performance (percentage of advanced students): historically underperforming schools (green line) versus all other schools (orange line); (b) rural or urban factors affecting academic performance: urban schools (blue line) versus rural schools (red line).

3.2. Academic Performance (Prediction versus Reality)

The authors evaluated five ML prediction methods, including decision tree, random forest, logistic regression, support vector machine, and neural network. Among all methods, the decision tree, random forest, logistic regression, and support vector machine achieved testing accuracy of 48%, 54%, 50%, and 51%, respectively (Table 3). The neural network achieved the highest 60% testing accuracy. As a result, this paper utilized the neural network method for the next step of the analysis.

When the authors applied the ML models, the prediction versus reality results were shown in Figure 6.

Table 3. Comparison of ML prediction method.

Method	Classifier	Training Accuracy	Testing Accuracy
Decision tree	DecisionTreeClassifier	94%	48%
Random forest	RandomForestClassifier	94%	54%
Logistic regression	LogisticRegression	48%	50%
Support vector machine	SupportVectorClassifier	59%	51%
Neural network	MLPClassifier	61%	60%



Figure 6. Percentage of advanced students by ML models (Prediction versus reality). Bold suggests the correctly predicted. The color green indicates a relatively high numerical value, with deeper shades of green indicating even higher values. Yellow signifies intermediate numerical values. Conversely, the color red indicates relatively low numerical values, with deeper shades of red indicating even lower values.

In the decision tree, of 8129 classes, 3904 classes were correctly predicted (48% of the total data, bold in Figure 6). For the prediction that was not completely correct, 3454 classes were predicted in the neighborhood group (42% of the total data). Ten percent of the predictions were far from the real situation.

In random forest, 4362 classes (54%) were correctly predicted. For the prediction that was not completely correct, 3198 classes (39%) were predicted in the neighborhood group. Seven percent of the predictions were far from the real situation.

In logistic regression, 4052 classes (50%) were correctly predicted. For the prediction that was not completely correct, 2984 classes (37%) were predicted in the neighborhood group. Thirteen percent of the predictions were far from the real situation.

In the support vector machine, 4162 classes (51%) were correctly predicted. For the prediction that was not completely correct, 2953 classes (36%) were predicted in the neighborhood group. 13% of the predictions were far from the real situation.

The most precise prediction in this paper came from neural networks: 4857 classes (60%) were correctly predicted. For the prediction that was not completely correct, 2896 classes (36%) were predicted in the neighborhood group. Only 4% of the predictions were far from the real situation. Moreover, neural networks can predict well for all groups (both good academic performance and bad academic performance). In sum, compared with other ML models, the neural network shows good prediction stability and accuracy.

When the authors compared the real percentage of advanced students versus neural network prediction (Figure 7c versus Figure 7d), the authors found there are only minor differences between the prediction and reality, and, as a result, the authors suggested that neural network was an accurate prediction method. The authors found that the neural network prediction can find the impacts of the county population (Figure 7a) and crime rate (Figure 7b). For example, the population and crime rate are high in Philadelphia county, the neural network obtained this information and predicted a significantly lower percentage of advanced students in Philadelphia county (Figure 7d): neural network prediction matched the real situation (Figure 7c) well.

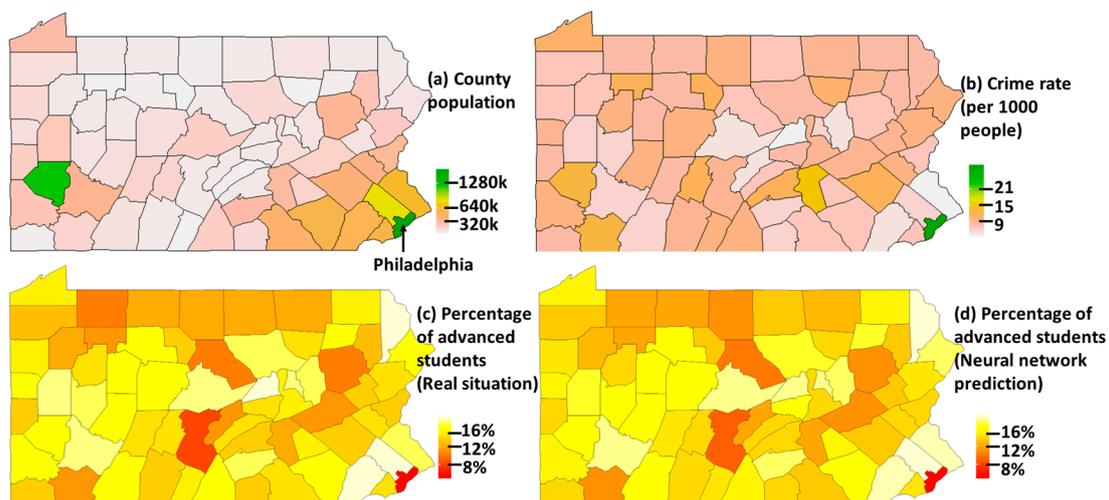


Figure 7. Heatmap of Pennsylvania showing (a) county population, (b) crime rate, (c) real percentage of advanced students, and (d) neural network predicted percentage of advanced students. Black arrow shows the special area of Philadelphia county.

4. Discussion

4.1. Academic Performance in Pennsylvania Schools

Some areas characterized by an inferior environment for learning (e.g., the city or county with a high population density and a considerable crime rate) could be harmful to many students, classes, and schools. The results could be explained by the less-educated people, drug abuse, poor security, gun issues, violence, and offenses. On the other hand,

well-educated people moved to a small county and created a positive learning atmosphere, better living conditions, and a supportive learning environment there, and the schools received good average academic performance. Despite this randomness of population quality within an area, the overall impact of an environment on academic performance is still recognizable. The results indicate that even with good education conditions (e.g., qualified teachers), large counties in Pennsylvania may not see a significant improvement in students' academic performance.

The reason may possibly be due to the fact that the areas with a high population density have an uneven demographic composition, varied population quality, less-educated parents, and the instabilities and uncertainties of potential offenders. Conversely, a small county with a healthy and harmonious cultural environment could contribute to raising students' academic performance in Pennsylvania schools. This could be due to the fact that these localities with well-educated people are safer and quieter and provide a healthy educational environment.

4.2. Advantages and Limitations of the ML Model

To predict academic performance, previous research normally used statistical models (e.g., Mplus or SPSS) to establish the relationships between such performance and SES, after which these were used in forecasting (Chang et al. 2020; Chen et al. 2021; Claver et al. 2020; Paulick et al. 2013). An example of the findings is the possibility that high income reduces parental stress, which may lead to a more stable study environment and improved academic performance (Owens 2018). The accuracy of classical methods hinges on the researchers' experience. By contrast, the current study introduced an ML model, which is simply a 'black box' that connects input (SES data) and output (academic performance) without considering relationships. The accuracy of this approach depends on data quality and quantity (Chen and Ding 2022). Through the ML method adopted in this work, academic performance in Pennsylvania's schools was successfully predicted.

The ML model is also encumbered with certain limitations, among which is its ineffectiveness in addressing the 'Black Swan' effect (Lorey et al. 2011). Most ML models generate results on the basis of data that were previously loaded into a computer program. If certain factors cannot be covered by a dataset, an ML model typically provides poor feedback. In our study, for example, for a school located in a high-density area with its surrounding environment suffering from a high crime rate, the ML model points to low academic performance. However, if this institution inputs an excellent teaching team and financial resources, it may achieve high academic performance. Moreover, there are some unaccountable factors that could not be explained by the datasets, which could cause miscalculations in an ML model.

4.3. Future Improvement of the ML Model

The future of ML models generally lies in two development directions: big data and novel algorithms. Considering that this is a relative feasibility analysis, it used only 33,870 records that span population, crime, and educational data (Considine and Zappalà 2002; Ginsburg and Bronstein 1993; Kurdek and Ronald 1988). In future research, if more related factors (e.g., family, economic, and transportation situations) can be considered in ML models, the authors believe such representations will generate more accurate predictions. At the same time, if data quantity can be improved (e.g., >100,000 records), more data can be used to support predictions. The availability of more data often relies on high-performance computing (Elsebakhi et al. 2015; Fox et al. 2019). If scientists in the future have access to better computers, they can also calculate more complex factors and larger amounts of data.

Algorithms are another component that can enhance ML models. In this study, five well-developed ML methods were compared: a decision tree, random forest, logistic regression, support vector machine, and neural networks. A neural network is one of the best models. The models recommended by the authors, including classification (Kotsiantis et al. 2006), KNN (Duivesteijn and Feelders 2008; Samworth 2012), linear discriminant

analysis (Izenman 2013; Xanthopoulos et al. 2013), K-means (Li et al. 2020; Likas et al. 2003), hidden Markov (Manogaran et al. 2018), and hierarchical planning (Mohr et al. 2018), can be explored by other researchers. These novel methods may also reduce computational requirements and increase predictive accuracy.

5. Conclusions

On the basis of big data (covering the Pennsylvania population, crime, and education data), this research demonstrated the feasibility of using ML models to predict class academic performance. To this end, the authors used an ML model that achieves fast and precise predictions: 60% of the predictions are accurate, 36% are highly close to reality, and only 4% exhibit substantial deviation from reality. This study confirmed that ML models are accurate and effective instruments. With ML models as grounding, the authors found that well-educated people in small counties that have lower crime rates could contribute to higher academic performance among Pennsylvania schools. Finally, SES exerts a significant impact on the rural–urban performance gap. The ML models are expected to provide assistance and guidance (e.g., decision making on issues that may affect performance, such as education budgets, hiring standards and practices, and teacher–student ratios) to education policymakers in the region in the future.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/socsci12030118/s1>, Figure S1: The feature importance of the factors; Figure S2: The correlation matrix of the samples; Table S1: Coding for machine learning and data analysis (Python); Table S2: Coding for feature importance (Python); Table S3: Coding for PA heatmap (R, Population, ×10,000); Table S4: Coding for PA heatmap (R, Total Offenses, ×1000); Table S5: Coding for PA heatmap (R, Real); Table S6: Coding for PA heatmap (R, Prediction); Table S7: Coding for PA heatmap (R, CrimeRate); Table S8: Tuning results by random search.

Author Contributions: Conceptualization, S.C. and Y.D.; methodology, Y.D.; software, Y.D.; validation, Y.D.; formal analysis, Y.D.; investigation, Y.D.; resources, S.C. and Y.D.; data curation, Y.D.; writing—original draft preparation, Y.D.; writing—review and editing, Y.D.; visualization, S.C.; supervision, Y.D.; project administration, Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors thank the knowledge and computation support from the School of Geography and the Environment, University of Oxford, United Kingdom.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Al-Jarrah, Omar, Paul Yoo, Sami Muhaidat, George Karagiannidis, and Kamal Taha. 2015. Efficient machine learning for big data: A review. *Big Data Research* 2: 87–93. [CrossRef]
- Alyahyan, Eyman, and Dilek Düşteğör. 2020. Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education* 17: 1–21. [CrossRef]
- Batrouni, Marwan, Aurélie Bertaux, and Christophe Nicolle. 2018. Scenario analysis, from BigData to black swan. *Computer Science Review* 28: 131–39. [CrossRef]
- Boxer, Paul, Grant Drawve, and Joel M Caplan. 2020. Neighborhood violent crime and academic performance: A geospatial analysis. *American Journal of Community Psychology* 65: 343–52. [CrossRef] [PubMed]
- Buenaño-Fernández, Diego, David Gil, and Sergio Luján-Mora. 2019. Application of machine learning in predicting performance for computer engineering students: A case study. *Sustainability* 11: 2833. [CrossRef]
- Bujang, Siti Dianah Abdul, Ali Selamat, Roliana Ibrahim, Ondrej Krejcar, Enrique Herrera-Viedma, Hamido Fujita, and Nor Azura Md. Ghani. 2021. Multiclass prediction model for student grade prediction using machine learning. *IEEE Access* 9: 95608–21. [CrossRef]
- Chang, Chi, Joseph Gardiner, Richard Houang, and Yan-Liang Yu. 2020. Comparing multiple statistical software for multiple-indicator, multiple-cause modeling: An application of gender disparity in adult cognitive functioning using MIDUS II dataset. *BMC Medical Research Methodology* 20: 275. [CrossRef]
- Chen, Shan, and Yuanzhao Ding. 2022. Machine Learning and Its Applications in Studying the Geographical Distribution of Ants. *Diversity* 14: 706. [CrossRef]

- Chen, Shan, Yuanzhao Ding, and Xin Liu. 2021. Development of the growth mindset scale: Evidence of structural validity, measurement model, direct and indirect effects in Chinese samples. *Current Psychology* 1–15. [\[CrossRef\]](#)
- Ciolacu, Monica, Ali Fallah Tehrani, Rick Beer, and Heribert Popp. 2017. Education 4.0—Fostering student’s performance with machine learning methods. Paper presented at 2017 IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME), Constanta, Romania, October 26–29, pp. 438–43.
- Claver, Fernando, Luis Manuel Martínez-Aranda, Manuel Conejero, and Alexander Gil-Arias. 2020. Motivation, discipline, and academic performance in physical education: A holistic approach from achievement goal and self-determination theories. *Frontiers in Psychology* 11: 1808. [\[CrossRef\]](#)
- Considine, Gillian, and Gianni Zappalà. 2002. The influence of social and economic disadvantage in the academic performance of school students in Australia. *Journal of Sociology* 38: 129–48. [\[CrossRef\]](#)
- Correa-Baena, Juan-Pablo, Kedar Hippalgaonkar, Jeroen van Duren, Shaffiq Jaffer, Vijay R. Chandrasekhar, Vladan Stevanovic, Cyrus Wadia, Supratik Guha, and Tonio Buonassisi. 2018. Accelerating materials development via automation, machine learning, and high-performance computing. *Joule* 2: 1410–20. [\[CrossRef\]](#)
- Duivesteyn, Wouter, and Ad Feelders. 2008. Nearest neighbour classification with monotonicity constraints. In *Machine Learning and Knowledge Discovery in Databases*. Berlin/Heidelberg: Springer, pp. 301–316.
- Ebel, Robert, and David Frisbie. 1972. *Essentials of Educational Measurement*. New Delhi: Prentice Hall of India, pp. 1–352.
- Elsebakhi, Emad, Frank Lee, Eric Schendel, Anwar Haque, Nagarajan Kathireason, Tushar Pathare, Najeed Syed, and Rashid Al-Ali. 2015. Large-scale machine learning based on functional networks for biomedical big data with high performance computing platforms. *Journal of Computational Science* 11: 69–81. [\[CrossRef\]](#)
- Fan, Xitao, and Michael Chen. 1998. Academic achievement of rural school students: A multi-year comparison with their peers in suburban and urban schools. *Journal of Research in Rural Education* 15: 31–46.
- Fedushko, Solomia, and Taras Ustyianovych. 2019. Predicting pupil’s successfulness factors using machine learning algorithms and mathematical modelling methods. In *Advances in Computer Science for Engineering and Education II*. Berlin/Heidelberg: Springer, pp. 625–36.
- Fox, Geoffrey, James Glazier, J. C. S. Kadupitiya, Vikram Jadhao, Minje Kim, Judy Qiu, James Sluka, Endre Somogyi, Madhav Marathe, Abhijin Adiga, and et al. 2019. Learning everywhere: Pervasive machine learning for effective high-performance computation. Paper presented at 2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Rio de Janeiro, Brazil, May 20–24, pp. 422–29.
- Ginsburg, Golda, and Phyllis Bronstein. 1993. Family factors related to children’s intrinsic/extrinsic motivational orientation and academic performance. *Child Development* 64: 1461–74. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hussain, Mushtaq, Wenhao Zhu, Wu Zhang, Syed Muhammad Raza Abidi, and Sadaqat Ali. 2019. Using machine learning to predict student difficulties from learning session data. *Artificial Intelligence Review* 52: 381–407. [\[CrossRef\]](#)
- Izenman, Alan Julian. 2013. Linear discriminant analysis. In *Modern Multivariate Statistical Techniques*. Berlin/Heidelberg: Springer, pp. 237–80.
- Jana, Strakova, Vladislav Tomasek, and Douglas Willms. 2006. Educational inequalities in the Czech Republic. *Prospects* 36: 517–27.
- Jung, Seok-Ki, and Tae-Woo Kim. 2016. New approach for the diagnosis of extractions with neural network machine learning. *American Journal of Orthodontics and Dentofacial Orthopedics* 149: 127–33. [\[CrossRef\]](#)
- Kemper, Lorenz, Gerrit Vorhoff, and Berthold U. Wigger. 2020. Predicting student dropout: A machine learning approach. *European Journal of Higher Education* 10: 28–47. [\[CrossRef\]](#)
- Kotsiantis, Sotiris, Ioannis Zaharakis, and Panagiotis Pintelas. 2006. Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review* 26: 159–90. [\[CrossRef\]](#)
- Kryst, Erica, Stephen Kotok, and Katerina Bodovski. 2015. Rural/urban disparities in science achievement in post-socialist countries: The evolving influence of socioeconomic status. *Global Education Review* 2: 60–77.
- Kurdek, Lawrence, and Sinclair Ronald. 1988. Relation of eighth graders’ family structure, gender, and family environment with academic performance and school behavior. *Journal of Educational Psychology* 80: 90. [\[CrossRef\]](#)
- Li, Liang, Jia Wang, and Xuetao Li. 2020. Efficiency analysis of machine learning intelligent investment based on K-means algorithm. *IEEE Access* 8: 147463–147470. [\[CrossRef\]](#)
- Likas, Aristidis, Nikos Vlassis, and Jakob J. Verbeek. 2003. The global k-means clustering algorithm. *Pattern Recognition* 36: 451–61. [\[CrossRef\]](#)
- Liu, Yanli, Yourong Wang, and Jian Zhang. 2012. New machine learning algorithm: Random forest. In *Information Computing and Applications*. Berlin/Heidelberg: Springer, pp. 246–52.
- Lorey, Johannes, Felix Naumann, Benedikt Forchhammer, Andrina Mascher, Peter Retzlaff, Armin ZamaniFarahani, Soeren Discher, Cindy Faehnrich, Stefan Lemme, Thorsten Papenbrock, and et al. 2011. Black swan: Augmenting statistics with event data. Paper presented at 20th ACM International Conference on Information and Knowledge Management, Glasgow, UK, October 24–28, pp. 2517–20.
- Lykourantzou, Ioanna, Ioannis Giannoukos, Vassilis Nikolopoulos, George Mpardis, and Vassili Loumos. 2009. Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education* 53: 950–65.

- Manogaran, Gunasekaran, Vijayakumar Varadarajan, Ramachandran Varatharajan, Priyan Malarvizhi Kumar, Revathi Sundarasekar, and Ching-Hsien Hsu. 2018. Machine learning based big data processing framework for cancer diagnosis using hidden Markov model and GM clustering. *Wireless Personal Communications* 102: 2099–116. [[CrossRef](#)]
- Mduma, Neema, Khamisi Kalegele, and Dina Machuve. 2019. A survey of machine learning approaches and techniques for student dropout prediction. *Data Science Journal* 18: 14. [[CrossRef](#)]
- Miller, Portia, Elizabeth Votruba-Drzal, and Rebekah Levine Coley. 2019. Poverty and academic achievement across the urban to rural landscape: Associations with community resources and stressors. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 5: 106–22. [[CrossRef](#)]
- Mohr, Felix, Marcel Wever, and Eyke Hüllermeier. 2018. ML-Plan: Automated machine learning via hierarchical planning. *Machine Learning* 107: 1495–515. [[CrossRef](#)]
- Musso, Mariel, Carlos Felipe Rodríguez Hernández, and Eduardo Cascallar. 2020. Predicting key educational outcomes in academic trajectories: A machine-learning approach. *Higher Education* 80: 875–94. [[CrossRef](#)]
- Owens, Ann. 2018. Income segregation between school districts and inequality in students' achievement. *Sociology of Education* 91: 1–27. [[CrossRef](#)]
- Papernot, Nicolas, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Berkay Celik, and Ananthram Swami. 2017. Practical black-box attacks against machine learning. Paper presented at 2017 ACM on Asia Conference on Computer and Communications Security, Abu Dhabi, United Arab Emirates, April 2–6, pp. 506–19.
- Paulick, Isabell, Rainer Watermann, and Matthias Nückles. 2013. Achievement goals and school achievement: The transition to different school tracks in secondary school. *Contemporary Educational Psychology* 38: 75–86. [[CrossRef](#)]
- Pedregosa, Fabian, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, and et al. 2011. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research* 12: 2825–30.
- Pomerat, John, Aviv Segev, and Rituparna Datta. 2019. On neural network activation functions and optimizers in relation to polynomial regression. Paper presented at 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, December 9–12, pp. 6183–85.
- Qazdar, Aimad, Brahim Er-Raha, Chihab Cherkaoui, and Driss Mammass. 2019. A machine learning algorithm framework for predicting students performance: A case study of baccalaureate students in Morocco. *Education and Information Technologies* 24: 3577–89. [[CrossRef](#)]
- Qi, Xinbo, Guofeng Chen, Yong Li, Xuan Cheng, and Changpeng Li. 2019. Applying neural-network-based machine learning to additive manufacturing: Current applications, challenges, and future perspectives. *Engineering* 5: 721–29. [[CrossRef](#)]
- Ramos, Raul, Juan Carlos Duque, and Sandra Nieto. 2012. Decomposing the rural-urban differential in student achievement in Colombia using PISA microdata. *SSRN Electronic Journal* 34: 379–412. [[CrossRef](#)]
- Rymarczyk, Tomasz, Edward Kozłowski, Grzegorz Kłosowski, and Konrad Niderla. 2019. Logistic regression for machine learning in process tomography. *Sensors* 19: 3400. [[CrossRef](#)]
- Samworth, Richard. 2012. Optimal weighted nearest neighbour classifiers. *The Annals of Statistics* 40: 2733–63. [[CrossRef](#)]
- Şara, Nicolae-Bogdan, Rasmus Halland, Christian Igel, and Stephen Alstrup. 2015. High-school dropout prediction using machine learning: A Danish large-scale study. Paper presented at 23rd European Symposium on Artificial Neural Networks, Bruges, Belgium, April 22–24; pp. 319–24.
- Sekeroglu, Boran, Kamil Dimililer, and Kubra Tuncal. 2019. Student performance prediction and classification using machine learning algorithms. Paper presented at 2019 8th International Conference on Educational and Information Technology, Cambridge, UK, March 2–4, pp. 7–11.
- Shakhovska, Natalya, Olena Vovk, Roman Hasko, and Yuriy Kryvenchuk. 2017. The method of big data processing for distance educational system. In *Advances in Intelligent Systems and Computing II*. Berlin/Heidelberg: Springer, pp. 461–73.
- Somvanshi, Madan, Pranjali Chavan, Shital Tambade, and Swati Shinde. 2016. A review of machine learning techniques using decision tree and support vector machine. Paper presented at 2016 International Conference on Computing Communication Control and automation (ICCUBEA), Pune, India, August 12–13, pp. 1–7.
- Willms, Douglas, Thomas Smith, Yanhong Zhang, and Lucia Tramonte. 2006. Raising and levelling the learning bar in central and Eastern Europe. *Prospects* 36: 411–18. [[CrossRef](#)]
- Xanthopoulos, Petros, Panos Pardalos, and Theodore Trafalis. 2013. Linear discriminant analysis. In *Robust Data Mining*. Berlin/Heidelberg: Springer, pp. 27–33.
- Yousafzai, Bashir Khan, Maqsood Hayat, and Sher Afzal. 2020. Application of machine learning and data mining in predicting the performance of intermediate and secondary education level student. *Education and Information Technologies* 25: 4677–97. [[CrossRef](#)]
- Zhou, Lina, Shimei Pan, Jianwu Wang, and Athanasios Vasilakos. 2017. Machine learning on big data: Opportunities and challenges. *Neurocomputing* 237: 350–61. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.