

Plagiarism - Report

Originality Assessment

27%



Overall Similarity

Date: Apr 20, 2024

Matches: 1818 / 4678 words

Sources: 13

Remarks: Moderate similarity detected, you may better improve the document (if needed).

Verify Report:

A Proportional Analysis Study on School Students' Performance System Using Machine Learning Algorithms

Ms. Moumita Sarkar, Assistance Professor, Department of Computational Sciences Brainware University, Kolkata, West Bengal. Email Moumita.kishalay@gmail.com

Dr.R.Naveenkumar, Associate Professor, ³ Department of Computer Science and Engineering Brainware University, Kolkata, West Bengal. Email: rnaveenkumarooty@gmail.com

Abstract

Objectives: Investigate global literacy rates and educational challenges, particularly focusing on gender disparities and the educational attainment of Sovereign states, while exploring the potential of Business Intelligence/Data Mining (BI/DM) techniques in addressing high school success. **Methods:** Utilized real-world data encompassing demographic, social, and academic characteristics, collected through certificates and questionnaires. Employed five-level/binary classification and regression tasks, testing four DM models (decision trees, random forest, neural networks, support vector machines) with and without leading notes. **Findings:** Comprehensive literacy rates were around 86% in 2016, with higher rates among younger age groups. Disparities in tertiary education attainment exist globally, with developed nations leading. Efforts to improve primary education access have shown progress but persisting challenges include gender disparities and insufficient resources. BI/DM techniques demonstrated potential in improving high school success rates, particularly in mathematics and Sovereign states. **Novelty:** This study highlights the innovative application of BI/DM techniques in addressing educational challenges, offering insights into improving high school success rates with real-world data and advanced modeling approaches.

Keywords: Business Intelligence, Classification, Regression, Decision Trees, Random Forest, ZeroR, Descriptive analysis.

1. Introduction

Education is a key factor for long-term economic progress. In the last decades the Sovereign states the level of education improved drastically[1][2]. However, the statistics holds Sovereign states thanks to its high level at the end of Europe student failure and dropout rates. For example, inside in 2006, the dropout rate in Sovereign states was 40% for people aged 18 to 24, while the European Union average in the figure was only 15% (Eurostat 2007)[3]. Failure in elementary mathematics and Sovereign states classes (native) is extremely bad because they provide the basic knowledge to be successful other school subjects (e.g.B. physics or history). **1 On the other hand,** the interest in business intelligence (BI)/Data Mining (DM) [14], derived from on the progress of information technology what to exponential corporate and organizational growth databases[4]. All this data contains valuable information, such as trends and patterns that can be used to improve decision making and optimize for success[1]. However, human experts are limited and may miss important points details. Therefore, the alternative is to use automated tools to analyze raw data and extract high-level information of interest to the decision maker. The education area provides fertile ground for BI applications as there are many data sources (e.g., traditional databases, websites) and various stakeholders (e.g.B. Students, teachers, administrative staff or graduates) [16]. For example, there are some interesting questions about this area responded using BI/DM techniques [15], [11] Who are the students who credit hours? Who is likely to return to the next lesson? What kind of courses can be offered to attract more students? What are the main reasons for changing students? Can student performance be predicted? Which are they Factors Influencing Student Achievement? This article focuses on the last two questions. modeling student Performance is an important tool for teachers and lecturer's students **because it can** help you understand better phenomenon and finally fix it. For example, Corrective action was taken by school employees for weak students (e.g., remedial courses). **As a result,** several studies have addressed similar issues. [16] used an associative principles approach to select a bad college student from Singapore for remedial education. Entry variables contained demographic characteristics (e.g., gender, region) and school performance in recent years' e in, the proposed solution passed the traditional

award procedure. [15] Online assessments by Michigan State University students were modeled using three assessment approaches (i.e., binary: pass/fail; 3 levels: low, medium, high; e.g., 9 points: from 1 - lowest score to 9 - highest score). THE Database contained 227 samples with online functionality (e.g., correct answers or homework attempts). The best results were achieved by the ranking team (e.g., decision tree and neural network) with an accuracy of 94% (binary), 72% (3 classes) and 62% (9 classes) classes).[8] used several DM algorithms to predict computational performance distance learning students at the university[5]. Different demographic physiognomies (e.g., gender, age, marital status) and leadership attributes (e.g., in the given task) were used as binary inputs pass/fail classifier. The best solution was obtained from A Naive Bayes method with 74% accuracy. That too found that past school years have much greater influence than the demographic variables. Recently, [9] collected data on **10 online tutoring system** for eighth **grade math tests** in the United States. The authors chose **a regression approach**, which was the goal to **predict math test score based on individual** ability. The authors used Bayesian networks and obtained the best result was a 15% prediction error. In this article we will analyze the latest **1 data from the** real world from two Sovereign states secondary schools. Two different sources were used: evaluation reports and questionnaires. Since the first residual information was included (e.g., To the right notes and number of absences available), incl was integrated by the latter, which allowed, among other things, a mix of different demographic, social, and educational characteristics (e.g., college age, alcohol use, maternal education). **The aim is to** anticipate the student results and, if possible, identify the key variables that influence school success/failure. two cores' classes (ex. Volume. Mathematics and Sovereign states) will be **based on the** three goals of the DM: i) binary classification (pass/fail); ii) Classification in five levels (excellent from I or excellent V - insufficient); AND iii) Regression, with numerical output between zero (0%) and twenty (100%). For each of these approaches, the three input configurations (e.g., with and without school grades) and four DM algorithms (e.g.B. decision trees, random forest) is tested. In addition, there will be an explanatory analysis must be performed on top models for identification purposes essential functions.

2. Material and methods

Student Data: It has been observed that most of the students go for admission the public and free education system. For marking a scale of grading point 20 is used where 0 is the lowest grade and 20 is highest. Students are evaluated **1** based on the scores in the three periods and the summation of these (G3 of Table4) corresponds to final score. This work has been done considering data collected during the school year 2005- 2006 from two public schools, from the Alen- tejo region of Sovereign states. Although there has been a trend for an increase of Information Technology. Figure 1 shows the snippet of the data.

Figure 1: Snippet of collected data

Investment from the Government, **the majority of** the Sovereign states public school information systems are very poor, relying mostly on paper sheets (which was the current case)[6][12]. **We designed the latter with closed questions (i.e., with predefined options) related to several demographic (e.g., mother's education, family income), social/emotional (e.g., alcohol consumption) (Pritchard and Wilson 2003) and school-related (e.g. number of past class failures) variables that were expected to affect student performance. The questionnaire was reviewed by school professionals and tested on a small set of 15 students in order to get feedback. The final version contained 37 questions in a single A4 sheet and it was answered in class by 788 students. Latter, 111 answers were discarded due to lack of identification details (necessary for merging with the school reports). 7 Finally, the data was integrated into two datasets related to Mathematics (with 395 examples) and the Sovereign states language (649 records) classes.**

2.1 Data Mining Models

Classification and regression- two crucial aspects and they require a supervised learning, where model is adjusted to a dataset made up of $k \in \{1, \dots, N\}$. The key difference is set **1** in terms of the production illustration. In classification, representations are habitually estimated using **the Percentage of Cor- react Classifications (PCC)**, while in regression the Root.

Table 1. Preprocessed student-related variables

Attribute Description (Domain)

sex student's sex (binary: female or male)

age student's age (numeric: from 15 to 22)

school student's school (binary: Gabriel Pereira or Mousinho da Silveira)

address student's home address type (binary: urban or rural)

Pstatus Parent's cohabitation status (binary: living together or apart)

Medu mother's education (numeric: from 0 to 4a)

Mjob mother's job (nominalb)

Fedu father's education (numeric: from 0 to 4a)

Fjob father's job (nominalb)

guardian student's guardian (nominal: mother, father or other)

famsize family size (binary: ≤ 3 or > 3)

famrel quality of family relationships (numeric: from 1 – very bad to 5 – excellent)

reason reason to choose this school (nominal: close to home, school reputation, course preference or other)

traveltime home to school travel time (numeric: 1 – < 15 min., 2 – 15 to 30 min., 3 – 30 min. to 1 hour or 4 – > 1 hour).

studytime weekly study time (numeric: 1 – < 2 hour, 2 – 2 to 5 hours, 3 – 5 to 10 hours or 4 – > 10 hours)

failures number of past class failures (numeric: n if $1 \leq n < 3$, else 4)

schoolsup extra educational school support 2 (binary: yes or no)

famsup family educational support (binary: yes or no)

activities extra-curricular activities (binary: yes or no)

paidclass extra paid classes (binary: yes or no)

internet Internet access at home (binary: yes or no)

nursery attended nursery school (binary: yes or no)

higher wants to take higher education (binary: yes or no)

romantic with a romantic relationship (binary: yes or no)

freetime free time after school (numeric: from 1 – very low to 5 – very high)

goout going out with friends (numeric: from 1 – very low to 5 – very high)

Walc weekend alcohol consumption (numeric: from 1 – very low to 5 – very high)

Dalc workday alcohol consumption (numeric: from 1 – very low to 5 – very high)

health current health status (numeric: from 1 – very bad to 5 – very good)

absences number of school absences (numeric: from 0 to 93)

G1 first period grade (numeric: from 0 to 20)

G2 second period grade (numeric: from 0 to 20)

G3 final grade (numeric: from 0 to 20)

0 – none, 1 – primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education.

teacher, health care related, civil services (e.g., administrative or police), at home or other.

Mean Squared (RMSE) is a popular metric (Witten and Frank 2005). ¹ A high PCC (i.e., near 100%) suggests a good classifier, while a regressor should present a low global error (i.e. RMSE close to zero). These metrics can be computed using the equations:

$$\Phi(i)=1,$$

$$\text{if } y_i = \hat{y}_i \text{ 0,}$$

$$\text{else } \sum \text{PCC} = N \Phi(i)/N \times 100 (\%)$$

$$i = q$$

$$1 \sum N$$

$$2 \text{ RMSE} = \sqrt{\sum_{i=1}^N (y_i - \hat{y}_i)^2 / N}$$

where \hat{y}_i denotes the predicted value for the i-th example.

In this work, the Mathematics and Sovereign state grades

(i.e., G3 of Table 1) will be modeled using three supervised approaches:

Binary classification – pass if $G3 \geq 10$, else fail;

5-Level classification – based on the Erasmus1 grade conversion system (Table 2);

Regression – the G3 value (numeric output between 0 and 20).

Figure 1 plots the respective histograms.

Several DM algorithms, each one with its own purpose and capabilities, have been proposed for classification and regression tasks. In this work, the grades will be modelled using three supervised approaches as followed by the pie chart. This pie representation can be turned into a set of IF-THEN rules that humans can understand. Breiman (2001) defined the Random Forest (RF) as an ensemble of T unpruned DT. Each tree is generated by averaging the outputs of the T trees and is based on a random feature selection from bootstrap training samples. When compared to a single DT, the RF is more difficult to read, but it is still possible to provide explanatory knowledge in terms of its input variable relevance. Nonlinear functions, such as Neural Networks (NN) and Support Vector Machines (SVM), have also been proposed for DM tasks, with better outcomes when nonlinearity is high [13]. The NN model in this work is based on the popular multilayer perceptron, with one hidden layer and H hidden nodes, whilst the SVM will utilize a gaussian kernel with one hyperparameter (γ). It should be highlighted that NN and SVM employ model representations that are challenging for humans to comprehend. Furthermore, because the DT/RF algorithms directly do internal feature selection, NN and SVM are more influenced by irrelevant inputs. Data has been classified as per the parameters shown in Table 2.

Table 2. Classification System

Mathematics

excellent/very good

good

satisfactory

sufficient

fail

Sovereign states/France

90 – 100

75 - 89

55 – 74

35 - 54

0 - 34

Sovereign states /France Out reach

80 – 100

60 - 80

45 – 60

35 - 44

0 - 34

Data

100

80

1 60

40

20

Construction of a branch representing a set of rules disguising values in a hierarchical form [10] which can be depicted by IF-ELSE paradigm to be understood by human easily. The Random Forest (RF) (Breiman 2001) is an ensemble of T unpruned DT. Each of these trees are based on a random feature selection from bootstrap training samples and the RF predictions are built by averaging the outputs of the T trees based on the knowledge of input variable. Nonlinear functions, such as Neural Networks (NN) and Support Vector Machines (SVM), have also been proposed for DM tasks [13], for getting better results when non-literacy is pretty high. We have used a popular multilayer perceptron-based NN model, having one hidden layer associated with H hidden nodes, but the SVM would use a Gaussian kernel with one hyper parameter (γ). Point to be noted that NN and SVM use model representations that are difficult for human being to understand, moreover NN and SVM are more affected by irrelevant inputs than the DT/RF algorithms since the latter explicitly performs an internal feature selection. Figure 2 shows student classification prediction system.

Figure 2.1: Sovereign States

Figure 2.2: Sovereign States Outreach

Figure 2.3: France Figure 2.4: France Outreach

Figure 2: Students Classification Prediction System

3. **1** Computational environment

All experiments reported in this study were conducted using the weka , an open-source library for the weka environment that facilitates the use of DM techniques (Figure 1). weka is a free and high-level matrix programming language with a powerful suite of tools for statistical and data analysis It can run in multiple platforms (e.g., Windows, MacOS or Linux) also betterment can be done in terms of new features by creating new packages. The weka library offers a set of coherent functions for classification and regression tasks. In particular, the library uses the rpart (DT), random Forest (RF),span class='highlighted color-1'>span> nnet (NN) and kernlab (SVM) packages. As an ex ample, the following WEKA was used during the DT predictive experiments of the Sovereign states binary classification.

3.1. Tool Used

Weka tool is a free data mining toolkit named after an inquisitive New Zealand bird, empowers researchers with a user-friendly interface and a rich set **11** of machine learning algorithms. It tackles **12** data preparation, classification, regression, clustering, and more, making it a valuable tool for data exploration and analysis.

Utilizing Weka's suite [11 of machine learning algorithms](#) within my research enabled me to comprehensively analyze my dataset. This facilitated the exploration of diverse approaches and ultimately led to the successful extraction of valuable insights. [1 Figure 3 shows the](#) environment of Weka tool.

Figure 3: The environment of the Weka tool

4. Result: Predictive Performance

[The NN and SVM models](#) required pre-processing before the models could be fitted. Nominal variables³ (e.g., Mjob) were converted to a 1-z-C coding [and all attributes were](#) normalized to a mean of zero [and one standard deviation](#) [13]. DM [models were fitted](#) later. Adjusted subdivision of DT node to reduce [sum of squares](#). For other [methods, the default](#) settings are RH (e.g., T=500), NN (e.g., G. [E = 100 epochs of BFGS algorithm](#)) and SVM (e.g., [Sequential Minimum Optimization Algorithm](#)). In addition, [the NN and SVM hyper parameters](#) [were optimized using an internal grid search](#) (e.g., Volume. only use [training data](#)) where $\epsilon \in \{0, 2, 4, 6, 8\}$ and $\gamma \in \{2-9, 2-7, 2-5, 2-3, 2-1\}$.

As per our suspicion grade [G1 and G2](#) would be having high impact we have tested [three input configurations for each DM model](#):

- [A - with all variables from Table 1 except G3 \(the output\)](#);
- [B - like A but without G2 \(the second period grade\)](#); and
- [C - like B but without G1 \(the first period grade\)](#).

[To access the](#) prediction results, 20 rounds of cross-validation ten times [13] were used for each configuration (200 simulations in total). In this scheme, the data [for a given run is randomly divided](#) into 10 equal subsets. Next, another subset [\(with 10% of the data\) is tested](#) [and the remaining data is used to](#) adjust [the DM technique](#). [At the end of this process, the test set](#) being evaluated contains the full data set, although 10 variants [of the same DM are used to create the predictions](#). A naive predictor (NV) is also tested [as a baseline comparison](#). For configuration A, this pattern [corresponds to the second period](#) class (G2 or binary/5-level versions). If the second rating [is not available](#) (configuration

B), the first periodic rating (or its binary/5-level variants) is used.

If no score is present (configuration C), the most common class (for classification tasks) or the average output (regression) is returned. The results of the test series are presented in Tables 3 to 5 as means and corresponding 95% confidence intervals for Student's t (Flexer 1996). As expected, the best performing configuration A is the great option for scientific categorization and regression based on the input choice A. This suggests that unevaluated inputs are useless in many of these cases. However, the scenario changes for the remaining experiments for that RF is the top choice in 8 events, followed by DT, which has the best 4 results. In general, nonlinear function methods (NN and SVM) are superior to tree methods. This behavior can be explained by the large number of irrelevant entries, as shown in the next section. Figure 3 represents the confusion matrices for the DT algorithm, the class of Sovereign states and configuration A. For the 5-level and binary classification, most of the values lie within/near the diagonal of the matrix., indicating a good fit. The figure also shows the HF scatterplot (predicted versus observed values) for the mathematical regression method. The HF predictions are over most of the 's output range ([5,20])

4.1 Descriptive Knowledge

The aim is not to deduce the predictive capabilities of every model, as per the measurement in the last section, but instead to give a modest explanation that recapitulates the best DM models. The entire dataset would be used in descriptive types of experiments. Furthermore, only the DT/RF algorithms will be considered, since it is easier to extract knowledge from these models had the best overall prediction results. Table 3 shows the summary of the testing process.

Test mode: 10-fold cross-validation

4 Classifier model (full training set)

ZeroR predicts class value:10.415189873417722

Time taken to build model: 0 seconds

Table 3. Testing procedure summary

Correlation coefficient

-0.1763

Mean absolute error

3.441

Root mean squared error

4.5925

Relative absolute error

100 %

Root relative squared error

100 %

Total Number of Instances

395

Test mode: evaluate on training data

=== Clustering model (full training set) ===EM==

Number of clusters selected by cross validation: 2

Number of iterations performed: 5

Table 4 shows **1** the relative importance (in percent) of each input variable as measured by the RF algorithm (Breiman 2001). To clarify the analysis, only the first five results are shown in the table. These five variables represent an overall impact between 43% and 77%, indicating many irrelevant inputs. As expected, student grades have a large influence (between 23% and 46%) in the models. suppose, G2 is the key feature for the A input choice sample, and G1 is extremely appropriate for the setup of B. Moreover, in relation with the performance of previous students the crucial fact is score is not available for any student so leads to failure. Total GP and Total MS as per gradation is shown in table 5.

Table 4. Relative importance (in percent) of each input variable as measured by the RF algorithm

Variable
Category
Mean
Std. Dev.
Total (GP)
Total (MS)
Course
Other
86.3105
60.6895
224.8695
18.689

Home

65.536

45.464

29.311

Reputation

74.1781

32.8219

Guardian

Mother

182.0001

92.9999

243.5585

Father

55.7695

36.2305

Other

6.7889

27.2111

Traveltime

Mean

1.3062

1.6716

Std. Dev.

0.5096

0.8711

Studytime

Mean

2.2075

1.7646

Std. Dev.

0.8546

0.7341

Failures

Mean

0.0037

0.8544

Std. Dev.

0.061

0.9857

Schoolsup

Yes

34.6416

18.3584

243.5585

No

208.917

137.083

Famsup

No

85.1267

69.8733

243.5585

Yes

158.4319

85.5681

Paid

No

113.5218

102.4782

243.5585

Yes

130.0368

52.9632

Activities

No

114.677

81.323

243.5585

Yes

128.8815

74.1185

Nursery

Yes

207.9538

108.0462

243.5585

No

35.6047

47.3953

MS

224.8695

18.689

Sex

F

141.1465

68.8535

243.5585

M

102.412

86.588

Age

Mean

16.4044

17.1556

Std. Dev.

1.1094

1.3778

Address

U

206.6688

102.3312

243.5585

R

36.8897

53.1103

Famsize

GT3

176.1999

106.8001

243.5585

LE3

67.3586

48.6414

Pstatus

A

27.9246

15.0754

243.5585

T

215.634

140.366

Medu

Mean

2.9918

2.3678

Std. Dev.

1.0212

1.0943

Fedu

Mean

2.7059

2.2312

Std. Dev.

1.0266

1.1152

Mjob

At Home

25.7086

35.2914

246.5585

Health

27.7073

8.2927

Other

86.4627

56.5373

Services

61.6379

43.3621

Teacher

45.042

14.958

Fjob

Teacher

22.5953

8.4047

246.5585

Other

134.1181

84.8819

Services

64.2975

48.7025

Health

13.9845

6.0155

At Home

11.5632

10.4368

Internet

Yes

208.5781

122.4219

243.5585

No

34.9804

33.0196

Romantic

No

170.6925

94.3075

243.5585

Yes

72.8661

61.1339

Famrel

Mean

4.0107

3.8398

Std. Dev.

0.8718

0.9219

Freetime

Mean

3.1532

3.3649

Std. Dev.

0.9747

1.0192

Dalc

Mean

1.1368

0.3436

Std. Dev.

2.0229

1.171

Walc

Mean

1.869

1.0046

Std. Dev.

2.9558

1.3958

Health

Mean

3.4377

1.4349

Std. Dev.

3.7382

1.2911

Absences

Mean

5.3414

8.3921

Std. Dev.

6.2874

7.2832

Table 5. GP and MS as per gradation

Grade

Category

Total (GP)

Total (MS)

G1

Mean

11.9645

9.247

Std. Dev.

3.097

2.945

G2

Mean

12.0341

8.6357

Std. Dev.

3.1432

3.7034

G3

Mean

12.0183

7.8915

Std. Dev.

3.6768

4.7116

ZeroR 5 model for the following dataset with an delicacy of attributes like 0&1.

0 251 (64%)

1 144 (36%)

1 Nevertheless, there are other relevant factors, such as school related (e.g., number of absences, extra school support or travel time), demographic (e.g., mother's work) and social (e.g., going out with friends, drinking alcohol). plots the best four DT. Again, the student grades are the most relevant features, appearing at the root of the trees, and only a small number (2 to 5) of the inputs considered are used. Some exciting rubrics can be mined from these trees, eg:

1. if $G1 < 10$ \square $G2 = 9$ \square $Mjob \in \{\text{teacher, other}\}$ then

pass;

2. if $G1 < 10$ \square $G2 = 9$ \square $Mjob \in \{\text{home, health, civilservices}\}$ then fail;

3. if $12 \leq G2 < 14$ \square $goout > 1$ then III;

4. if $12 \leq G2 < 14$ \square $goout = 1$ then II;

5. if 11 \square $absences < 7$ then 13;

6. if 11 \square $absences \geq 7$ then 11;

These rules show the influence of the mother's job (rules 1–2), going out with friends (rules 3–4) and

the number of absences (rules 5–6). In ZeroR algorithm for example the value represents the easiest method **3** for classification which is based on target and does not take any other predictions into account. The ZeroR classifier's just predicting the majority of a given category class. **5** Although there is no predictability power in ZeroR, it is useful for determining a baseline performance as a benchmark for other classification methods.

Time taken to build model (full training data): 0.55 seconds.

5. Conclusion

Proper education is the pivotal element behind the improvement of any society and the role **1** of Business Intelligence (BI)/Data Mining (DM) techniques to filter out an augmented acquaintance from the original raw data cannot be denied. Studies are still going on for the improvement of quality of education based on these techniques. In this paper, we attempted to expect the grades of students of the secondary level in two years of basic school (math and Sovereign states) based totally on grades from preceding years (1st and second grade), demographic statistics, social facts, and other records based on four DM methods like Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM) and Neural Networks (NN). A high degree of prediction accuracy is possible if the grades are known. This confirms the conclusion of [7] Study success is strongly influenced by previous success. This work relied on offline learning as DM techniques were applied after data collection. Additional features and valuable feedback can be collected by using an automated online learning environment which use student prediction engine in the backend. To enrich the student database, we wish to expand our trials to more students of different grade level with varying methods of automatic feature selection. specifically, nonlinear characteristic strategies (e.g. B. NN and SVM), that are touchier to irrelevant enter. similarly studies (e.g., sociological studies) is also had to understand why and the way sure variables (e.g., reasons for college preference, dad and mom' profession, or alcohol intake) have an effect on student overall performance. And also we had analyzed ZeroR algorithm for using diverse attributes as an example for various element with various time slot and it had been shown the automated result for the predicted results.

6. References

1. Han, J.; Kamber, M.; Pei, J. [13 Data Mining: Concepts and Techniques](#) (3rd ed.). Elsevier Science & Technology: Amsterdam, The Netherlands, 2012.
2. Mathew, S.; Abraham, J.T.; Kalayathankal, S.J. Data mining techniques and methodologies. *Int. J. Civ. Eng. Technol.* 2018, 9, 246–252.
3. Smale-Jacobse, [9 A.E.](#); Meijer, A.; Helms-Lorenz, M.; Maulana, R. [Differentiated Instruction in Secondary Education: A Systematic Review of Research Evidence.](#) *Front. Psychol.* 2019, 10, 2366 [css\[PubMed\]](#).
4. Mosa, M.A. Analyze students' academic performance [3 using machine learning](#) techniques. *J. King Abdulaziz Univ. Comput. Inf. Technol. Sci.* 2021, 10, 97–121.
5. [Nasir, M.U.](#); [Khan, S.](#); [Mehmood, S.](#); [Khan, M.A.](#); [Rahman, A.-U.](#); [Hwang, S.O.](#) [8 IoMT-Based Osteosarcoma Cancer Detection in Histopathology Images Using Transfer Learning Empowered with Blockchain, Fog Computing, and Edge Computing.](#) *Sensors* 2022, 22, 5444.
6. [6 Nasir, M.U.](#); [Zubair, M.](#); [Ghazal, T.M.](#); [Khan, M.F.](#); [Ahmad, M.](#); [Rahman, A.-U.](#); [Al Hamadi, H.](#); [Khan, M.A.](#); [Mansoor, W.](#) [Kidney Cancer Prediction Empowered with Blockchain Security Using Transfer Learning.](#) *Sensors* 2022, 22, 7483.
7. [Rahman, A.-U.](#); [3 Alqahtani, A.](#); [Aldhaffer, N.](#); [Nasir, M.U.](#); [Khan, M.F.](#); [Khan, M.A.](#); [Mosavi, A.](#) [Histopathologic Oral Cancer Prediction Using Oral Squamous Cell Carcinoma Biopsy Empowered with Transfer Learning.](#) *Sensors* 2022, 22, 3833.
8. [Farooq, M.S.](#); [Abbas, S.](#); [Rahman, A.U.](#); [Sultan, K.](#); [Khan, M.A.](#); [Mosavi, A.](#) [A Fused Machine Learning Approach for Intrusion Detection System.](#) *Comput. Mater. Contin.* 2023, 74, 2607–2623.
9. [Rahman, A.U.](#); [Dash, S.](#); [Luhach, A.K.](#); [Chilamkurti, N.](#); [Baek, S.](#); [Nam, Y.](#) A Neuro-fuzzy approach for user behaviour classification and prediction. *J. Cloud Comput.* 2019, 8, 17.
10. [1 Breiman, L.](#); [Friedman, J.](#); [Olshen, R.](#); [Stone, C.](#) [Classification and Regression Trees.](#) *Wadsworth Publishing Company*: Belmont, CA, USA, 1984.
11. [Cortez, P.](#) [RMiner: Data Mining with Neural Networks and Support Vector Machines using R.](#) In [Introduction to Advanced Scientific Softwares and Toolboxes](#), edited by [Rajesh, R.](#), Springer,

Singapore, 2010.

12. Flexer, A. Statistical Evaluation of Neural Networks Experiments: Minimum Requirements and Current Practice. In Proceedings of the 13th European Meeting on Cybernetics and Systems Research, Vienna, Austria, vol. 2, pp. 1005-1008, 1996.

13. Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Verlag: New York, USA, 2001.

14. Kotsiantis, S.; Pierrakeas, C.; Pintelas, P. Predicting Students' Performance in Distance Learning Using Machine Learning Techniques. Applied Artificial Intelligence, vol. 18, no. 5, pp. 411-426, 2004.

15. Luan, J. Data Mining and Its Applications in Higher Education. New Directions for Institutional Research, vol. 113, pp. 17-36, 2002.

16. Ma, Y.; Liu, B.; Wong, C.; Yu, P.; Lee, S. Targeting the right students using data mining. In Proc. Of . of 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Boston, USA, 457-464.

Sources

1	https://www.academia.edu/3114409/Using_data_mining_to_predict_secondary_school_student_performance INTERNET 18%
2	https://archive.ics.uci.edu/dataset/320/student_performance INTERNET 3%
3	https://www.mdpi.com/1424-8220/22/14/5444 INTERNET 1%
4	https://stackoverflow.com/questions/24425728/learning-weka-precision-and-recall-wiki-example-to-arff-file INTERNET 1%
5	https://saedsayad.com/zeror.htm INTERNET 1%
6	https://www.mdpi.com/1424-8220/22/19/7483/notes INTERNET <1%
7	https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1078&context=ny_pubs INTERNET <1%
8	https://www.researchgate.net/publication/362158594_IoMT-Based_Osteosarcoma_Cancer_Detection_in_Histopathology_Images_Using_Transfer_Learning_Empowered_with_Blockchain_Fog_Computing_and_Edge_Computing INTERNET <1%
9	https://www.researchgate.net/publication/337451346_Differentiated_Instruction_in_Secondary_Education_A_Systematic_Review_of_Research_Evidence/fulltext/5dd88b42458515dc2f459179/Differentiated-Instruction-in-Secondary-Education-A-Systematic-Review-of-Research-Evidence.pdf INTERNET <1%
10	http://www3.dsi.uminho.pt/pcortez/student.pdf INTERNET <1%
11	https://sourceforge.net/projects/weka/ INTERNET <1%
12	https://www.udemy.com/course/weka-data-mining-with-open-source-machine-learning-tool/ INTERNET <1%
13	https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques INTERNET <1%

EXCLUDE CUSTOM MATCHES

OFF

EXCLUDE QUOTES ON

EXCLUDE BIBLIOGRAPHY ON