

Article

Analysis of the Factors Affecting Student Performance Using a Neuro-Fuzzy Approach

Mahmoud Abou Naaj , Riyadh Mehdi , Elfadil A. Mohamed  and Mirna Nachouki * 

Information Technology Department, Ajman University, Ajman 346, United Arab Emirates

* Correspondence: mirna@ajman.ac.ae

Abstract: Predicting students' academic performance and the factors that significantly influence it can improve students' completion and graduation rates, as well as reduce attrition rates. In this study, we examine the factors influencing student academic achievement. A fuzzy-neural approach is adopted to build a model that predicts and explains variations in course grades among students, based on course category, student course attendance rate, gender, high-school grade, school type, grade point average (GPA), and course delivery mode as input predictors. The neuro-fuzzy system was used because of its ability to implicitly capture the functional form between the dependent variable and input predictors. Our results indicate that the most significant predictors of course grades are student GPA, followed by course category. Using sensitivity analysis, student attendance was determined to be the most significant factor explaining the variations in course grades, followed by GPA, with course delivery mode ranked third. Our findings also indicate that a hybrid course delivery mode has positively impacted course grades as opposed to online or face-to-face course delivery alone.

Keywords: educational data mining; neuro-fuzzy systems; student performance; prediction; neural networks; fuzzy systems



Citation: Abou Naaj, M.; Mehdi, R.; Mohamed, E.A.; Nachouki, M. Analysis of the Factors Affecting Student Performance Using a Neuro-Fuzzy Approach. *Educ. Sci.* **2023**, *13*, 313. <https://doi.org/10.3390/educsci13030313>

Academic Editor: James Albright

Received: 21 January 2023

Revised: 8 March 2023

Accepted: 15 March 2023

Published: 17 March 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

Higher education institutions are always searching for ways to improve the academic performance of their students. In the last two decades, a considerable amount of research has been conducted to predict students' performance and obtain the most critical factors affecting their academic performance [1–3]. Various techniques have been used to predict students' academic achievement, ranging from conventional approaches to the most modern educational data mining methodologies.

Due to the increased digitization of academic institutions, educational data mining (EDM) can more easily extract meaningful information from their datasets [1]. EDM includes fields of study such as automatic student performance discrimination, instructional deficiency analysis, student adaptive learning ability analysis, and other related topics [4,5]. Due to the lack of significant technological advancements at the time, researchers previously only utilized data mining techniques for education in the student performance prediction domain in a crude manner, with a small number of applications. However, a revolution in educational technology has recently been prompted by sophisticated internet technology, and various data mining techniques have been used to advance the development of educational digitization [6].

Many researchers have recently applied EDM methods to examine students' performance in the academic curriculum based on their score data. For example, [7] used a data mining technique to analyze students' learning habits and suggest processes to improve their performance. In addition, [8] used the k-means clustering technique to forecast the understanding level of students in a specific course based on their test scores and provide pertinent information for preparing final examination questions.

Researchers have long studied the factors that affect students' performance at higher education institutions. Among the factors that have been studied and found to significantly impact students' performance are their attendance, type of secondary certificate, grade of secondary school, course grade, student gender, and parent educational level [3]. The correlation between attendance and academic performance is examined in a study by [9]. In their investigation, more than 900 individuals enrolled in a bachelor's degree program were included. They discovered that attendance boosts academic performance, and that this benefit persists at higher grade percentiles.

Other research investigating the impact of students' attendance on their performance can be observed in [2]. In their study, the authors determined that there is an interconnectivity between the attendance of classes and subsequent academic performance. Various factors might have a considerable impact on students' performance. Therefore, there is a need to investigate which among the abovementioned factors has the greatest impact on student performance.

2. Objectives and Scope of the Study

The primary objective of this study is to build a predictive as well as an explanatory model of student performance at the course level. Specifically, we aim to use readily available input attributes to (a) determine the relative importance of each input variable in predicting student performance, (b) identify the relative importance of the input attributes as factors explaining variations in course grades, (c) understand how the COVID-19 pandemic affected student performance due to the change in course delivery mode from face-to-face to fully online and subsequently, to hybrid learning, and (d) learn whether course attendance has any effect on student performance. Most studies in the literature attempt to predict student performance based on a set of a variety of attributes representing academic and non-academic factors; the novelty of our work is that in addition to identifying the most important predictors of student performance, our approach also addresses the problem of determining the strength of the causal relationship between course grades and the various input variables used.

The attributes considered in this study as input predictors are attendance rate, course category, gender, high school grade, school type, grade point average (GPA), and delivery mode as input predictors. The high school grade is used as a proxy for a host of demographic factors that can affect student performance. The delivery mode input variable represents the teaching methodology adopted before, during, and after the COVID-19 pandemic, in which teaching was conducted fully online during the pandemic.

Although there are numerous machine learning techniques that can be used, to build our model, we have opted for a fuzzy-neural approach that predicts course grades based on the above-mentioned attributes and examines the relative predictive power of each input predictor based on observations of changes in the root mean square error (RMSE) when a specific input is temporarily dropped from the model. The model was then used to run sensitivity analyses to quantitatively determine the relative significance of each input variable in explaining course grade variations among students. The neuro-fuzzy system was used because of its ability to implicitly capture the functional form between the dependent variable and input predictors [10–12], as well as the experience of the authors with using this machine-learning technology.

3. Related Work

Predicting university students' performance is a hot research topic that has attracted the attention of a considerable number of researchers and academics. As a result, numerous methods are used to predict students' performance. These methods range from traditional statistical methods to recent machine learning and data mining techniques.

3.1. Predicting Students' Performance Using Machine Learning Methods

Machine learning (ML) explores how a computer can learn based on large data [13]. ML can extract implicit information from data to uncover trends and connections that human reviewers would have missed. ML, which can be used to predict outcomes, has various uses in many application areas. ML is a viable method used for predicting students' academic performance. In recent years, the research area of educational data mining (EDM) has increased in popularity. EDM uses several fields, including computer science, education, and statistics, and ML, to address complex problems in teaching and learning [14,15].

In EDM techniques, forecasting student performance is performed using predictive modeling. Several methods are used to build predictive modeling: classification, regression, and categorization. Classification is the most commonly used technique to predict students' academic achievement. Several studies have been proposed that use ML techniques: for example, ref. [16] applied the ID3 decision tree induction technique to build students' academic achievement prediction models on data regarding female students enrolled in the bachelor's program in the Information Technology (IT) department, King Saud University, Riyadh, Saudi Arabia. The results demonstrate that precise predictions can be made based on the students' performance in second-year courses. In the same direction, ref. [17] developed a model using various ML methods, such as random forest, nearest neighbor, support vector machines, and logistic regression, to anticipate final exam marks in undergraduate courses based on midterm exam grades using naive Bayes and k-nearest neighbor algorithms. With just three inputs—midterm test grades, departmental data, and faculty data—the proposed model was able to attain a classification accuracy between 70 and 75%.

The paragraphs below provide an overview of the most common ML techniques used to predict students' academic achievement. The approaches include neural networks, decision trees, random forests, naive Bayes, k-nearest neighbors, support vector machines, Petri nets, and generalized nets.

Artificial Neural network (ANN) is a set of input/output nodes with weighted connections [13]. The main steps involved in building a model using ANN are preparing the data, training the ANN, testing the model results, and deploying the model. During the learning stage, the network learns by modifying the weights so as to be able to predict the class value correctly. Neural networks are prevalent in EDM. They can detect all the possible relationships between predictor variables. They exploit nonlinear associations underlying a dataset to achieve classification. A typical neural network comprises several layers of nodes and uses a mathematical function in each node to contribute to the classification decision. Weights are applied to the input in each layer between the input and output layers. Usually, the goal of these weights is to reduce the number of classification errors. The classification layer is the output layer. Neural networks perform well for categorical and continuous variables and scale to large sample numbers. In [18], the authors conducted a study with 162,030 students from both public and private universities in Colombia to classify how well students performed at the university. The results suggest that artificial neural networks (ANN) can be used systematically, with an accuracy of 82%. Additionally, the authors found that ANNs perform better than traditional machine-learning techniques for evaluation measures such as recall and the F1 score. Furthermore, their research demonstrated that ANN-based predictive models outperform other predictive approaches in terms of accuracy and suitability for handling classification issues with unbalanced data. The authors in [19] found that ANNs outperform conventional traditional prediction techniques in accurately forecasting students' academic performance, according to a review of 21 papers on the topic. The application of neural networks in education continues, with a number of recent studies being conducted; for example, ref. [20] used a feedforward spike neural network to forecast students' academic achievements. The results of their study show that the prediction accuracy of student performance approaches 70.8% using the proposed model. Such results justify the viability of ML techniques in predicting students' academic performance. To forecast students' performance at higher education institutions utilizing transitory online education systems, ref. [21] created three models based on probabilistic neural networks,

support vector machines, and discriminant analysis. This study enabled the organization to identify students who were expected to succeed and those who might fail.

Decision trees are a popular prediction technique because of their ease and directness in recognizing the structure of small or large data and predicting their value. A typical decision tree consists of a binary tree, with thresholds at each node that determine which node needs to be explored next. The leaf or terminal node regulates the classification of a case. This technique helps identify and visualize the various variables that influence an outcome and is frequently accurate and computationally fast for small to medium datasets. It provides a valuable explanation, since a tree's top nodes contain the most critical variables in the classification process. However, in large datasets, interpretability and accuracy are reduced. Several algorithms were used to build decision tree models, such as the iterative dichotomiser (ID3) and classification and regression trees (CART) [13]. Most decision tree algorithms start with a training set of records and their connected class labels. The training dataset will be recursively divided into smaller subsets as the tree is built. In the next step, the built tree will be pruned. In [22], the authors reported that decision trees have the highest accuracy rate for predicting students' academic performance. Furthermore, the authors of [23] researched students' academic performance using a decision tree algorithm with parameters including a student's educational information and activities. They collected data on 22 students at the undergraduate level at Oman's private higher education institution in the spring 2017 semester. The results show that the random forest tree algorithm showed better accuracy than the comparative decision tree algorithms.

Random forests are well-known techniques used for classification and regression. The algorithm builds multiple distinct classifiers and then groups them to make a final prediction. According to [24], random forests create bootstrapped datasets from an initial training dataset for training. Instead of considering all the variables at each node, the method obtains the optimal split condition for a random subset of the variables. This randomness produces more diverse trees, improving the overall performance by significantly aggregating multiple unrelated trees and reducing the estimator's variance. Before starting the training step, the random forest algorithm should be provided with the values of the three parameters, which are node size, the number of trees, and the number of features sampled. The final prediction is performed by aggregating the class predicted by each tree, using a simple majority vote in classification problems.

Naive Bayes is also a simple technique used to make a prediction. Depending on the assumption that all x values are independent, two probabilities—the probability of each class (y) and the conditional probability (y given x)—are calculated [25]. The naïve Bayesian classifiers presume that the effect of an attribute value on a given class is independent of the values of the other attributes [13]. As explained in [13], assume that there are m classes, given a record, X , the classifier will forecast that X belongs to the class having the highest posterior probability, conditioned on X . That is, the naïve Bayesian classifier predicts that tuple X belongs to the class C_i if and only if

$$P(C_i | X) > P(C_j | X) \text{ for } 1 \leq j \leq m, j \neq i$$

This technique produces accurate results with large training sets and numerous predictor variables. In addition, it is practical when categorical predictor variables are present in multi-class classification situations.

K-nearest neighbors is a method used for predicting new cases by summarizing the k-nearest cases. Instead of using the training and test sets, it uses the entire dataset. The neighbor cases are identified using the Euclidean distance [26]. As described in [13], k-nearest-neighbor classifiers learn by proximity by comparing a given test record with training records that are similar to it. The training records are defined by n features. Each record represents a point in an n -dimensional space. Using this approach, all the training records are stored in n -dimensional pattern space. In case the tuple is unknown, a k-nearest-neighbor classifier explores the pattern space for the k-training records which are nearest to the unknown record. These k-training records are called the k-nearest neighbors of

the unidentified records. This technique often requires reducing the dimensions, as it is difficult to measure the distance with large numbers of dimensions [27].

Support vector machines (SVMs) are supervised learning methods for classification. SVMs are highly accurate modeling techniques. The SVM uses nonlinear mapping to convert the original training data into a higher dimension. The SVM technique separates the data into classes by obtaining the hyperplane, where larger margins are optimal. The SVM finds this hyperplane using support vectors, and it works well when the data are linearly separable and linearly inseparable [13]. The SVM is a binary ML with some highly sophisticated properties. Unlike neural networks, SVM training always discovers a global solution. The SVM is slower to train than other methods, but is very efficient at classifying new data [28]. Moreover, unlike regression techniques and naive Bayes, this method needs to make stronger assumptions. Due to regularization, overfitting usually does not occur.

Hot techniques and approaches used for predicting students' performance continue to emerge. Petri nets, a well-known mathematical and graphical technique with the advantages of graphical notation and simple semantics, have recently been used to forecast students' grades based on their usage of the Learning Management System [29]. Generalized nets (GNs), an extension of Petri nets, are among the recent approaches that have gained popularity and are used for predicting students' performance and finding dependencies between various criteria [30]. It has more and better modeling capabilities than several other types of Petri nets.

3.2. Predicting Students' Performance Using Statistical Methods

Traditional techniques, such as statistical methods, have long been studied and used to predict academic performance and identify factors that significantly impact students' performance. For example, ref. [2] used quantitative data and descriptive analyses to determine a link between student attendance in class and student performance in examinations.

Another example of using a statistical method to present a correlation between student attendance and academic performance is documented in [31], which used Spearman's correlation coefficient, a nonparametric statistical test, to determine the correlation between student attendance and academic performance.

3.3. Predictive Attributes for Student Performance

The primary focus of many researchers is determining which variables impact a student's academic performance and developing a predictive model. Currently, researchers understand that a student's academic success is influenced by their family, home, demographics, school, and environment. Student demographics and socioeconomic factors include place of birth, disability, educational background, employment history of parents, region of residence, gender, socioeconomic index, health insurance, frequency of socializing with friends, and financial status [32,33]. Pre-enrollment factors include high school or grade point average, entry requirements, SAT scores, IELTS scores [34], and physics and math grades; enrollment factors include enrollment date, enrollment test results [35], the number of courses the students had previously taken, the type of study program, and the mode of study [36,37]; tertiary academic factors include attendance, the number of assessment submissions, student engagement, major, and the amount of time left to complete the degree [32,37,38], and the LMS-based data have all been studied in previous analyses regarding the prediction of student academic performance.

Academic performance has been examined from various angles, including knowledge score, grade point average (GPA), course grades [39], semester or final results of a student, as well as the graduate or dropout status of a student, which were frequently used as categorical variables in [32,33,37,40–42]. These are regarded as critical predictors of academic potential. However, the CGPA and GPA are the most popular metrics used by researchers to assess students' academic success.

Various aspects, including demographic, educational, social, and family backgrounds, are incorporated into a classifier model for predicting academic achievement [40]. Demo-

graphic variables include gender, age, race, and location of residence; educational attributes include quizzes, grades for assignments and midterms, attendance rates, and study techniques. Lifestyle, time spent on social media, the number of close friends, etc., are a few categories of social traits. The number of kids, parents' earnings, and educational backgrounds are all part of the family background. Additionally, it is highlighted that parents' social or economic status impacts their children's academic performance and examination grade position, either favorably or unfavorably [43,44].

Students' performance is impacted by various factors, although these factors vary from person to person and organization to organization. For example, the authors of [45] claim multiple connections and exchanges between teachers and students due to courses and the perceived value connected to their academic success. In addition, multiple literature reviews indicate that environmental, economic, social, and psychological variables significantly impact students' academic achievement [46]. In [47], the authors integrated the association between students' location-based, semester-wise, behavioral, and academic aspects using a geospatial-based machine learning technique. Past performance, social standing, and the semester were some of the variables identified. Students' academic success has also been influenced by additional elements, such as the importance of the course experience, effort, motivation, and learning methodologies [48].

4. Research Methodology

In order to develop models, the ANFIS analytical methodology combines the strengths of fuzzy logic with those of neural networks [10,11]. Neural networks estimate the parameters of the model and the representation of information. Contrastingly, fuzzy inference techniques are more akin to human reasoning and enhance the model's capability to deal with uncertainty [10,11]. ANFIS learns the characteristics of a particular pattern through the examples submitted to the system. Prediction is enhanced by progressively changing the system's parameters until they converge to the error criterion set by the system. In order to avoid making any assumptions about the complexity, ambiguity, functional form, or other characteristics of the causal relationship between the course grade and its determinants, an ANFIS-based course grade prediction model was developed [49]. For traditional techniques, such as multilinear regression (MLR) [49], the form of the functional relationship between the output and input predictors must be known beforehand.

A fuzzy model is a mathematical representation of ambiguity and imprecise information; these models can recognize, represent, manipulate, interpret, and utilize ambiguous and uncertain facts and information. There are two types of fuzzy inference models, which differ mainly in how various types of information are represented. "Linguistic models", known as Mamdani fuzzy models, are included in the first category. These are built using fuzzy reasoning and sets of fuzzy if-then rules, with vague predicates [50]. The Takagi-Sugeno inference process serves as the foundation for other classes of fuzzy models. These models have logical rules, with fuzzy premises and consequents. Linguistic models' capacity to express qualitative and quantitative information is assimilated into fuzzy models based on the Takagi-Sugeno reasoning process [12]. The membership functions of the Sugeno output have linear or constant functional forms, which is the fundamental functional distinction between the Mamdani and Sugeno systems. When dealing with noisy input data, a Sugeno FIS outperforms a Mamdani FIS in terms of computational effectiveness, projected accuracy, and robustness [51,52]. A Sugeno system's output function is guaranteed by the adaptive approaches used to calculate the parameters of the membership functions, which are better suited for quantitative analysis [51]. As a result, this study used the ANFIS approach, based on the Sugeno FIS model, to predict student course grades. The following subsections address the model's architecture and learning procedures.

4.1. Adaptive Neuro-Fuzzy Inference System

Fuzzy logic systems require an external system to set and modify their settings, since they lack the capacity for learning and the adaptability of neural networks. Although they

can learn from an input, neural networks are opaque because the weights of the connections between their neurons store the information used in their reasoning [53]. The adaptive neuro-fuzzy inference system (ANFIS) is an intelligent system architecture that combines the advantages of neural networks and fuzzy logic principles in a single framework. ANFIS is an artificial neural network based on the Takagi–Sugeno inference system [10]. The function of the inference system is to produce an approximative output, and the task of the neural network component is to estimate the fuzzy rules’ membership function parameters from the input and output samples [10,11]. Figure 1 illustrates the ANFIS architecture.

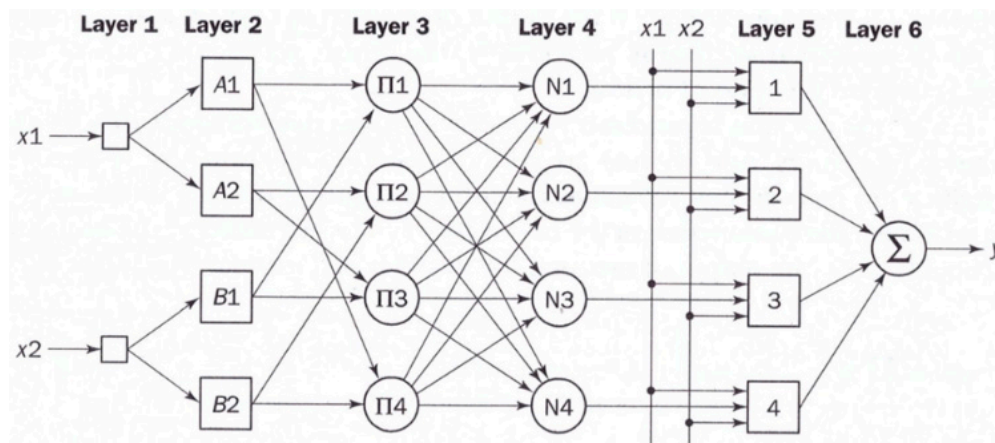


Figure 1. The adaptive Sugeno neuro-fuzzy inference system architecture.

A typical architecture of ANFIS is shown in Figure 1 [11]. A circle describes a fixed node, whereas a square indicates an adaptive node. For a first-order Sugeno fuzzy model, a two-rule rule base is expressed as follows:

1. If x is A_1 and y is B_1 , then $f_1 = p_1x + q_1y + r_1$
2. If x is A_2 and y is B_2 , then $f_2 = p_2x + q_2y + r_2$

Assume that the membership functions of fuzzy sets A_i, B_i for $i = 1, 2$ are given as μ_{A_i}, μ_{B_i} . In this work, the authors used Gaussian membership functions,

$$\mu_{A_i}(x) = \frac{1}{1 + \left(\frac{x-c_i}{a_i}\right)^{2b_i}} \tag{2}$$

A product T-norm (logical and) is chosen to evaluate the rules. Assess the rule premises results in,

$$w_i = \mu_{A_i}(x)\mu_{B_i}(y), i = 1, 2. \tag{3}$$

Evaluating the implication and the rule consequences gives,

$$f(x, y) = \frac{w_1(x, y)f_1(x, y) + w_2(x, y)f_2(x, y)}{w_1(x, y) + w_2(x, y)} \tag{4}$$

Leaving the arguments out,

$$f = \frac{w_1f_1 + w_2f_2}{w_1 + w_2} \tag{5}$$

The above equation is rewritten as,

$$f = \bar{w}_1f_1 + \bar{w}_2f_2 \tag{6}$$

where,

$$\bar{w}_i = \frac{w_i}{w_1 + w_2}$$

The Sugeno-type neuro-fuzzy system was implemented using MATLAB programming toolboxes.

4.2. Anfis Learning

Membership function parameters are learned with ANFIS by applying a gradient descent technique and a least-squares error criterion [10]. Each neuron's membership function is initially assigned an activation function, and the centers of the membership functions are adjusted so that each of their widths and slopes properly overlap and equally cover the entire input range [11]. Training is conducted in two steps: a forward and backward pass for each epoch. Each rule's subsequent parameters are calculated in the forward pass using a least squares technique. The network can calculate the output error once the parameters of the rule's consequents are determined. Using the errors as inputs for the backward pass, the back-propagation learning algorithm modifies the parameters of the membership functions [11]. The antecedent and consequent parameters of Jang's learning algorithm are optimized during the learning process [10]. The subsequent parameters are changed in the forward pass, but the antecedent parameters are left unchanged. On the other hand, the backward pass modifies the antecedent parameters, while leaving the consequent parameters unchanged. When the training dataset is limited, human experts can also choose the membership functions and their parameters and maintain them during the training process [11].

4.3. Dataset

The course-grade (CG) dataset was constructed from transcripts of 650 students studying for the degrees of BSc in Information Technology, Information Systems, and Computer Engineering, collected from Spring 2013 to Spring 2022, at the authors' institution. The course grade target variable was measured on a scale of 0–100. Courses with pass-fail grades, withdrawn status, and transferred from other institutions were ignored. The input variables affecting course grades examined in this study were: course category, student course attendance percentage, gender, high-school grade, school type, grade point average (GPA), and delivery mode. Each course was assigned to one of eight categories, and certificates issued by high schools were classified into five categories based on curricula. The delivery mode input variable had three categorical values: face-to-face, online, and hybrid. This variable was used to examine the consequences of moving from face-to-face learning to fully online course delivery during the COVID-19 pandemic and the subsequent return to on-campus learning, and assessment complemented by online delivery. Categorical variables were converted into numerical types using MATLAB tools. Table 1 shows the values of the categorical input variables used in the study. The total number of records comprising the dataset was 15,596, provided by the registration office.

Table 1. Categorical input variables and corresponding values.

| Categorical Input | Values |
|-------------------|--|
| course category | Programming, Mathematics, Core Information Technology, Advanced Information Technology, Advanced Information Systems Courses, Engineering, General Education, and Business Courses |
| gender | Male, Female |
| school type | National High School Certificate, American High School Certificate or equivalent, British GCE High School Certificate or equivalent, Pakistani/Indian High School Certificate, and African/Iranian High School Certificate |
| delivery mode | Face-to-face, online, hybrid |

4.4. Predictive and Explanatory Model Performance

The process of designing, implementing, and training a numerical or a machine-learning algorithm using an appropriate dataset for forecasting new or future observations is known as predictive modeling. The main objective of predictive modelling is to predict the outcome of a new observation from its input values [54]. The degree of the underlying causal relationship between the input and output variables is reflected in the model's explanatory component. However, it may not always reflect the model's capacity for prediction [55]. Because the emphasis is on the association rather than the causation between the input variables and the dependent variable, it is not necessary for predictive modeling to ascertain the role of each input variable in the underlying causal structure [55]. The metric computation measuring each input variable's relative importance in predicting or explaining the target variable uses different datasets [56]. As a result, the prediction accuracy of metrics produced from the data used to train and build the model is frequently overstated. As a result, the testing data subset provides a more accurate context for assessing a model's ability to predict [57]. Measures such as the RMSE on testing data not included in the model's construction are used to evaluate the predictive power of input variables. On the other hand, the metrics that are used to evaluate an input variable's explanatory power are calculated by assessing how the model's performance varies as a result of introducing variations in the input variable's values in the training dataset [56].

5. Results and Discussion

The following subsections discuss the model's effectiveness as a tool for predicting and explaining variations in course grades.

5.1. Model Performance

The CG dataset was divided into two subsets for training and testing, with training making up 70% of the whole. To avoid overfitting, the model was trained using k-fold cross-validation with $k = 5$. An RMSE of 9.3 was observed in the model's values predicted using testing data. As a result, the model's accuracy for the test data was 90.62%. Figure 2 shows that the model captures the functional link between the course grade as a target variable and the input factors looked at in this study. The model is then used to observe how each input variable affected the determination of the course grade in the subsequent steps.

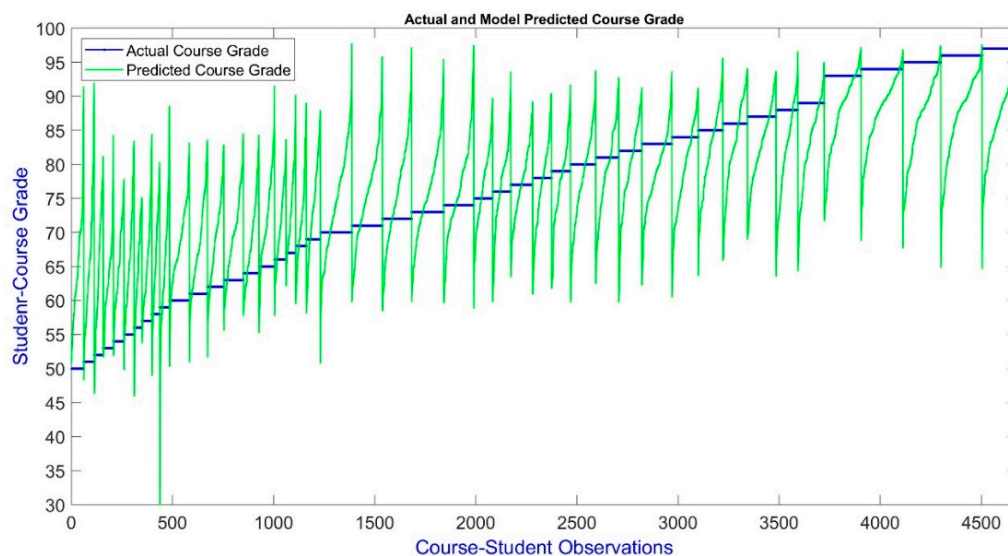


Figure 2. The actual and predicted course grades.

5.2. Predictive Importance of Input Variables

Backward elimination is a method that is frequently used to evaluate the relative prediction power of an input variable [58]. All input variables were initially used to

train the ANFIS model, and the appropriate RMSE was calculated using the testing data. Subsequently, the model was retrained using the same testing data after eliminating each input variable one at a time and observing the change in RMSE. The percentage of the predictive influence of a given input predictor is inversely correlated with the magnitude of the change in the RMSE that results from the removal of that predictor [58]. In Figure 3, the RMSE that results from eliminating one input variable at a time as a predictor is shown as the value on the accompanying horizontal bar. Contrast this value with the RMSE value shown on the top bar (no input variable removed). The change quantifies how much the RMSE has increased due to dropping that input as a predictor, demonstrating its predictive power. As can be observed in Figure 3, GPA is the most important predictor, followed by course category. The effects of the remaining predictors are almost equal.

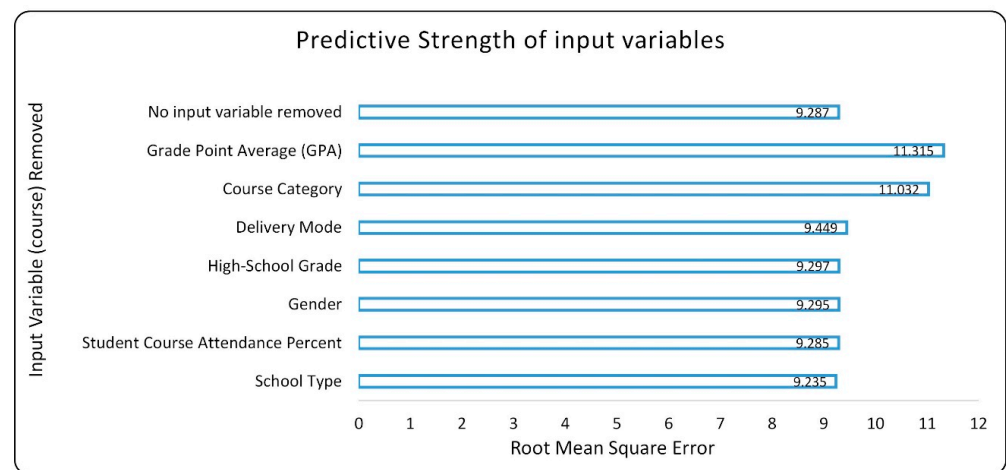


Figure 3. The relative predictive power of each input variable.

5.3. Model Explanatory Performance

We performed a sensitivity analysis where several approaches were presented for neural network-based models [59] to determine the causal influence of each input variable on the course grade. It was demonstrated that the partial derivative and input perturbation algorithms outperformed previously used methods [60]. However, the partial derivative approach has two significant flaws. First, neural networks with non-differentiable activation functions cannot be evaluated using this method. Second, it might be challenging to determine precisely how much a change in a particular input variable affects the output [61]. For this study, we opted for the perturbation strategy. This method introduces noise into one input variable, while leaving the remaining inputs unaltered. Calculations were performed to determine the percentage change in the output variable resulting from a certain perturbation level for each input variable individually. The process was repeated for various levels of disturbance. The input variable causing the highest percentage change provided the best explanation for variations in the system's output [62]. Training data were used to determine the values of the sensitivity spectra at increasing levels of perturbation, ranging from 0 to 20% in steps of 0.01 after the neuro-fuzzy model had been trained. Figure 4, which shows the output of the sensitivity index for each input variable, indicates that attendance is the primary factor influencing course grades, followed by GPA. As shown in Figure 4, the other input components are presented in a distant third place, with little variation between them.

During the COVID-19 pandemic, courses were delivered fully online from the Spring 2020 to the Spring 2021 semesters. One of the objectives of this study is to understand how the COVID-19 pandemic affected student performance due to the change in course delivery mode from face-to-face, to fully online, and subsequently, to hybrid learning after the pandemic. Using a *t*-test analysis, we observed a statistical difference at the 99% confidence level ($\alpha = 0.01$) in students' performance as reflected in the average grade of all courses

presented in Table 2. Our interpretation is that during the online course delivery, students were obliged to engage and take an active role in their own learning, which may have led to this slight improvement. Moreover, this engagement seems to have continued after the pandemic and the return to on-campus teaching, supplemented with online delivery and assessment, which enhanced their performance even further. The delivery mode attribute ranked third in terms of explaining variations in course grades.

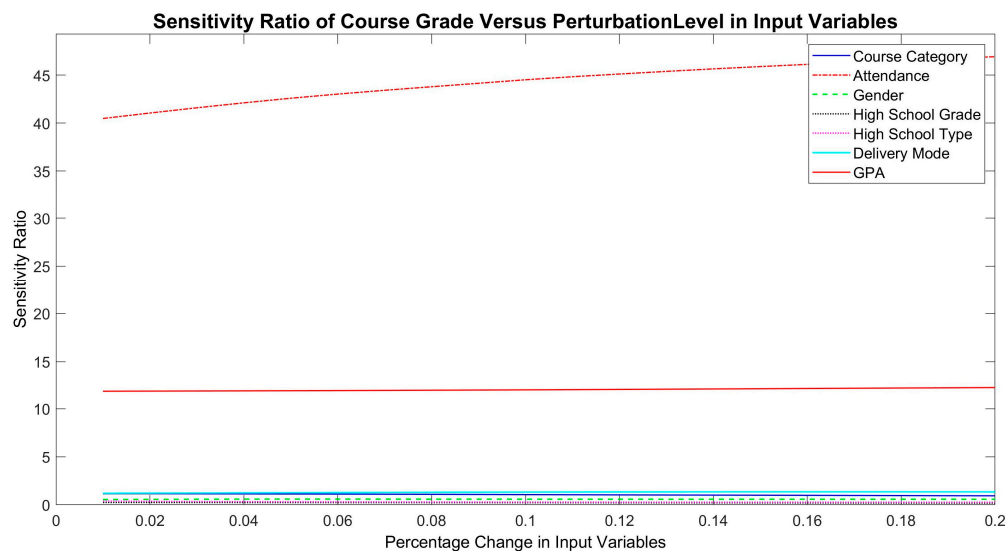


Figure 4. Sensitivity ratio values with respect to perturbation levels in the input variables.

Table 2. Average course grades before, during, and after the pandemic period.

| Delivery Mode | Mean Course Grade | Standard Deviation | Sample Size |
|---------------|-------------------|--------------------|-------------|
| Face-to-face | 75.28 | 14.17 | 5815 |
| Fully online | 76.85 | 12.73 | 3636 |
| Hybrid | 78.84 | 12.28 | 6145 |

5.4. Limitations and Model Validity

Our model established the importance of the numerous variables that contributed to the prediction of course grades and explained the factors that impact them. However, the model’s efficacy could be affected by a few issues. First, according to [62], grades may be impacted by various grading criteria and methods, as well as variations in the instructors’ distribution of accepted grades. Therefore, a high faculty teaching course turnout may affect the model’s capacity to predict and explain these outcomes. Second, changes in the course material, instructional methodologies, and assessment methods might result from reviews of academic programs and affect grades. Third, the impact of curricular modifications on introductory computer programming courses was examined in [63]. They discovered that the selection and sequencing of curriculum courses, the time spent in class, the proportion of lectures to practical sessions, and the class load all impacted students’ academic achievement. Because of this, an estimate of a student’s grade might not be accurate if these variables change throughout the program.

6. Conclusions

In this study, course category, student course attendance rate, gender, high school grade, school type, grade point average (GPA), and delivery mode were employed as input predictors of student course performance. Our results show that the two most significant predictors of course grade are student grade point average and course category, in that order. However, the remaining input variables seem to have no predictive power. Sensitivity analysis was incorporated into the model to examine the significance of each input variable

in explaining variations in course grades. Our findings indicate that student attendance is the most crucial variable explaining variations in course grades, followed by student grade point average. We also determined that course delivery positively impacted course grades and ranks a distant third in its importance as a factor in determining course grades. In conclusion, our ANFIS computational model can shed light on the predictive power of the input variables used, as well as determine those attributes that explain variations in course grades. This work will serve as a benchmark for future work that will examine other factors that may influence course grades, such as scores on English fluency tests and some demographic factors, including age and place of residence. Other methodologies, such as random forests, Petri net models, and the InterCriteria analysis approach based on the notion of intuitionistic fuzzy sets, will be investigated for consistency with this work's findings.

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

Funding: This research received no funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to institutional restrictions.

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

References

1. Al Breiki, B.; Zaki, N.; Mohamed, E.A. Using educational data mining techniques to predict student performance. In Proceedings of the 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA), Ras Al Khaimah, United Arab Emirates, 19–21 November 2019; pp. 1–5.
2. Doniņa, A.; Svētiņa, K.; Svētiņš, K. Class Attendance As a Factor Affecting Academic Performance. In Proceedings of the International Scientific Conference, Rezekne, Latvia, 20 May 2020; Volume 6, pp. 578–594.
3. Etemadpour, R.; Zhu, Y.; Zhao, Q.; Hu, Y.; Chen, B.; Sharier, M.A.; Zheng, S.; SPAiva, J.G. Role of absence in academic success: An analysis using visualization tools. *Smart Learn. Environ.* **2020**, *7*, 2. [[CrossRef](#)]
4. Injadat, M.; Moubayed, A.; Nassif, A.B.; Shami, A. Multi-split optimized bagging ensemble model selection for multi-class educational data mining. *Appl. Intell.* **2020**, *50*, 4506–4528. [[CrossRef](#)]
5. Livieris, I.E.; Drakopoulou, K.; Tampakas, V.T.; Mikropoulos, T.A.; Pintelas, P. Predicting secondary school students' performance utilizing a semi-supervised learning approach. *J. Educ. Comput. Res.* **2019**, *57*, 448–470. [[CrossRef](#)]
6. Amrieh, E.A.; Hamtini, T.; Aljarah, I. Mining educational data to predict student's academic performance using ensemble methods. *Int. J. Database Theory Appl.* **2016**, *9*, 119–136. [[CrossRef](#)]
7. El-Halees, A. Mining students data to analyze learning behavior: A case study. In Proceedings of the 2008 International Arab Conference of Information Technology (ACIT2008)—Conference Proceedings, Hammamet, Tunisia, 16–18 December 2008; p. 137.
8. Ayesha, S.; Mustafa, T.; Sattar, A.; Khan, I. Data mining model for higher education system. *Eur. J. Sci. Res.* **2018**, *43*, 24–29.
9. Karnik, A.; Kishore, P.; Meraj, M. Examining the linkage between class attendance at university and academic performance in an International Branch Campus setting. *Res. Comp. Int. Educ.* **2020**, *15*, 371–390. [[CrossRef](#)]
10. Jang, J.S. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 665–685. [[CrossRef](#)]
11. Negnevitsky, M.; Intelligence, A. A guide to intelligent systems. In *Artificial Intelligence*; Prentice Hall: Harlow, UK, 2017.
12. Rutkowski, L. Fuzzy Inference Systems. In *Flexible Neuro-Fuzzy Systems: Structures, Learning and Performance Evaluation, The International Series in Engineering and Computer Science*; Springer: Boston, MA, USA, 2004; Volume 771, pp. 27–50.
13. Han, J.; Kamber, M.; Pei, J. *Data Mining Concepts and Techniques*, 3rd ed.; University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University: Burnaby, BC, Canada, 2012.
14. Du, X.; Yang, J.; Hung, J.L.; Shelton, B. Educational data mining: A systematic review of research and emerging trends. *Inf. Discov. Deliv.* **2020**, *48*, 225–236. [[CrossRef](#)]
15. Anjewierden, A.A.; Kolloffel, B.; Hulshof, C. Towards educational data mining. Using data mining methods for automated chat analysis and support inquiry learning processes. In Proceedings of the International Workshop on Applying Data Mining in e-Learning, Crete, Greece, 17–18 September 2007; Available online: <https://core.ac.uk/display/20962888> (accessed on 16 August 2021).

16. Altujjar, Y.; Altamimi, W.; Al-Turaiki, I.; Al-Razgan, M. Predicting critical courses affecting students performance: A case study. *Procedia Comput. Sci.* **2016**, *82*, 65–71. [CrossRef]
17. Yağcı, M. Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learn. Environ.* **2022**, *9*, 11. [CrossRef]
18. Rodríguez-Hernández, C.F.; Musso, M.; Kyndt, E.; Cascallar, E. Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Comput. Educ. Artif. Intell.* **2021**, *2*, 100018. [CrossRef]
19. Baashar, Y.; Alkaws, G.; Mustafa, A.; Alkahtani, A.A.; Alsariera, Y.A.; Ali, A.Q.; Hashim, W.; Tiong, S.K. Toward predicting student's academic performance using artificial neural networks (ANNs). *Appl. Sci.* **2022**, *12*, 1289. [CrossRef]
20. Liu, C.; Wang, H.; Yuan, Z. A Method for Predicting the Academic Performances of College Students Based on Education System Data. *Mathematics* **2022**, *10*, 3737. [CrossRef]
21. Cazarez, R.L. Accuracy comparison between statistical and computational classifiers applied for predicting student performance in online higher education. *Educ. Inf. Technol.* **2022**, *27*, 11565–11590. [CrossRef]
22. Chaka, C. Educational Data Mining, Student Academic Performance Prediction, Prediction Methods, Algorithms and Tools: An Overview of Reviews. Available online: <https://www.preprints.org/manuscript/202108.0345/v1> (accessed on 16 August 2021).
23. Hasan, R.; Palaniappan, S.; Raziff, A.R.; Mahmood, S.; Sarker, K.U. Student academic performance prediction by using decision tree algorithm. In Proceedings of the 2018 4th International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, Malaysia, 13 August 2018; pp. 1–5.
24. Díaz-Uriarte, R.; Alvarez de Andrés, S. Gene selection and classification of microarray data using random forest. *BMC Bioinform.* **2006**, *7*, 3. [CrossRef] [PubMed]
25. Rish, I. An empirical study of the naive Bayes classifier. In Proceedings of the IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, Seattle, WA, USA, 4 August 2001; Volume 3, pp. 41–46.
26. Zhang, Z. Introduction to machine learning: K-nearest neighbors. *Ann. Transl. Med.* **2016**, *4*, 218. [CrossRef] [PubMed]
27. Beyer, K.; Goldstein, J.; Ramakrishnan, R.; Shaft, U. When is "nearest neighbor" meaningful? In Proceedings of the International Conference on Database Theory, Jerusalem, Israel, 10 January 1999; Springer: Berlin/Heidelberg, Germany, 1999; pp. 217–235.
28. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, UK, 2000.
29. Balogh, Z.; Kuchárik, M. Predicting student grades based on their usage of LMS moodle using Petri nets. *Appl. Sci.* **2019**, *9*, 4211. [CrossRef]
30. Atanassov, K.; Sotirova, E.; Andonov, V. Generalized net model of multicriteria decision making procedure using intercriteria analysis. In *Advances in Fuzzy Logic and Technology 2017: Proceedings of: EUSFLAT-2017–The 10th Conference of the European Society for Fuzzy Logic and Technology, September 11–15, 2017, Warsaw, Poland IWIFSGN'2017–The Sixteenth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets, September 13–15, 2017, Warsaw, Poland, Volume 1 10*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 99–111.
31. Kassarnig, V.; Bjerre-Nielsen, A.; Mones, E.; Lehmann, S.; Lassen, D.D. Class attendance, peer similarity, and academic performance in a large field study. *PLoS ONE* **2017**, *12*, e0187078. [CrossRef]
32. Imran, M.; Latif, S.; Mehmood, D.; Shah, M. Student academic performance prediction using supervised learning techniques. *Int. J. Emerg. Technol. Learn.* **2019**, *14*, 92–104. [CrossRef]
33. Zeineddine, H.; Braendle, U.; Farah, A. Enhancing prediction of student success: Automated machine learning approach. *Comput. Electr. Eng.* **2021**, *89*, 106903. [CrossRef]
34. Fateh ALLAH, A.Q. Using Machine Learning to Support Students' Academic Decisions. Ph.D. Thesis, The British University in Dubai (BUiD), Dubai, United Arab Emirates, 2019.
35. Mengash, H.A. Using data mining techniques to predict student performance to support decision-making in university admission systems. *IEEE Access* **2020**, *8*, 55462–55470. [CrossRef]
36. Berens, J.; Schneider, K.; Görtz, S.; Oster, S.; Burghoff, J. Early Detection of Students at Risk—Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods. CESifo Working Paper No. 7259. 2018. Available online: <https://ssrn.com/abstract=3275433> (accessed on 16 August 2021). [CrossRef]
37. Kemper, L.; Vorhoff, G.; Wigger, B.U. Predicting student dropout: A machine learning approach. *Eur. J. High. Educ.* **2020**, *10*, 28–47. [CrossRef]
38. Xu, J.; Moon, K.H.; Mvd, S. A machine learning approach for tracking and predicting student performance in degree programs. *IEEE J. Sel. Top. Signal. Process.* **2017**, *11*, 742–753. [CrossRef]
39. Nabil, A.; Seyam, M.; Abou-Elfetouh, A. Prediction of students' academic performance based on courses' grades using deep neural networks. *IEEE Access* **2021**, *9*, 140731–140746. [CrossRef]
40. Poudyal, S.; Mohammadi-Aragh, M.J.; Ball, J.E. Prediction of Student Academic Performance Using a Hybrid 2D CNN Model. *Electronics* **2022**, *11*, 1005. [CrossRef]
41. Mehdi, R.; Nachouki, M. A neuro-fuzzy model for predicting and analyzing student graduation performance in computing programs. *Educ. Inform. Technol.* **2022**, *28*, 1–30. [CrossRef]
42. Nachouki, M.; Abou Naaj, M. Predicting Student Performance to Improve Academic Advising Using the Random Forest Algorithm. *Int. J. Distance Educ. Technol. (IJDET)* **2022**, *20*, 1–7. [CrossRef]

43. Vermunt, J.D. Relations between student learning patterns and personal and contextual factors and academic performance. *High. Educ.* **2005**, *49*, 205–234. [[CrossRef](#)]
44. Azhar, M.; Nadeem, S.; Naz, F.; Perveen, F.; Sameen, A. Impact of parental education and socioeconomic status on academic achievements of university students. *Eur. J. Psychol. Res.* **2014**, *1*, 1–9.
45. Tsinidou, M.; Gerogiannis, V.; Fitsilis, P. Evaluation of the factors that determine quality in higher education: An empirical study. *Qual. Assur. Educ.* **2010**, *18*, 227–244. [[CrossRef](#)]
46. You, J.W. Testing the three-way interaction effect of academic stress, academic self-efficacy, and task value on persistence in learning among Korean college students. *High. Educ.* **2018**, *76*, 921–935. [[CrossRef](#)]
47. Musaddiq, M.H.; Sarfraz, M.S.; Shafi, N.; Maqsood, R.; Azam, A.; Ahmad, M. Predicting the Impact of Academic Key Factors and Spatial Behaviors on Students' Performance. *Appl. Sci.* **2022**, *12*, 10112. [[CrossRef](#)]
48. Diseth, Å.; Pallesen, S.; Brunborg, G.S.; Larsen, S. Academic achievement among first semester undergraduate psychology students: The role of course experience, effort, motives and learning strategies. *High. Educ.* **2010**, *59*, 335–352. [[CrossRef](#)]
49. Pal, M.; Bharati, P. Introduction to correlation and linear regression analysis. In *Applications of Regression Techniques*; Springer: Singapore, 2019; pp. 1–18.
50. Tanaka KSugeno, M. Introduction to fuzzy modelling. In *Fuzzy Systems: Modeling and Control*; Nguyen, H.T., Sugeno, M., Eds.; Kluwer: New York, NY, USA, 1998; pp. 63–89.
51. Subhedar, M.; Birajdar, G. Comparison of mamdani and sugeno inference systems for dynamic spectrum allocation in cognitive radio networks. *Wirel. Pers. Commun.* **2013**, *71*, 805–819. [[CrossRef](#)]
52. Mitra, S.; Hayashi, Y. Neuro-fuzzy rule generation: Survey in soft computing framework. *IEEE Trans. Neural Netw.* **2000**, *11*, 748–768. [[CrossRef](#)] [[PubMed](#)]
53. Geisser, S. *Predictive Inference: An Introduction*; Chapman and Hall/CRC: New York, NY, USA, 2017.
54. Shmueli, G. To explain or to predict? *Stat. Sci.* **2010**, *25*, 289–310. [[CrossRef](#)]
55. Geisser, S. The predictive sample reuse method with applications. *J. Am. Stat. Assoc.* **1975**, *70*, 320–328. [[CrossRef](#)]
56. Mosteller FTukey, J.W. *Data Analysis and Regression*; Addison-Wesley: Reading, MA, USA, 1977.
57. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *97*, 273–324. [[CrossRef](#)]
58. Cao, M.; Alkayem, N.F.; Pan, L.; Novák, D.; Rosa, J.L. Advanced methods in neural networks-based sensitivity analysis with their applications in civil engineering. In *Artificial Neural Networks-Models and Applications*; IntechOpen: London, UK, 2016; pp. 335–353. [[CrossRef](#)]
59. Wang, W.; Jones, P.; Partridge, D. Assessing the impact of input features in a feedforward neural network. *Neural Comput. Appl.* **2000**, *9*, 101–112. [[CrossRef](#)]
60. Cheng, A.Y.; Yeung, D.S. Sensitivity analysis of neocognitron. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **1999**, *29*, 238–249. [[CrossRef](#)]
61. Lamy, D. Modelling and sensitivity analysis of neural network. *Math. Comput. Simul.* **1996**, *40*, 535–548. [[CrossRef](#)]
62. Tomkin, J.H.; West, M.; Herman, G.L. An improved grade point average, with applications to C.S. undergraduate education analytics. *ACM Trans. Comput. Educ. (TOCE)* **2018**, *18*, 1–6. [[CrossRef](#)]
63. e Silva, I.H.; Pacheco, O.; Tavares, J. Effects of curriculum adjustments on first-year programming courses: Students performance and achievement. In *Proceedings of the Frontiers in Education Conference, Boulder, CO, USA, 5–8 November 2003*; Volume 1, p. T4C-10.

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.