



# Article Predicting the Impact of Academic Key Factors and Spatial Behaviors on Students' Performance

Muhammad Hammad Musaddiq <sup>(D)</sup>, Muhammad Shahzad Sarfraz, Numan Shafi, Rabia Maqsood, Awais Azam and Muhammad Ahmad \*

Department of Computer Science, National University of Computer and Emerging Sciences (NUCES), CFD Campus, Chiniot 35400, Punjab, Pakistan

\* Correspondence: mahmad00@gmail.com

Abstract: Quality education is necessary as it provides the basis for equality in society. It is also significantly important that educational institutes be focused on tracking and improving the academic performance of each student. Thus, it is important to identify the key factors (i.e., diverse backgrounds, behaviors, etc.) that help students perform well. However, the increasing number of students makes it challenging and leaves a negative impact on credibility and resources due to the high dropout rates. Researchers tend to work on a variety of statistical and machine learning techniques for predicting student performance without giving much importance to their spatial and behavioral factors. Therefore, there is a need to develop a method that considers weighted key factors which have an impact on their performance. To achieve this, we first surveyed by considering experts' opinions in selecting weighted key factors using the Fuzzy Delphi Method (FDM). Secondly, a geospatial-based machine learning technique was developed which integrated the relationship between students' location-based features, semester-wise behavioral features, and academic features. Three different experiments were conducted to prove the superiority and predict student performance. The experimental results reveal that Long Short-Term Memory (LSTM) achieved higher accuracy of 90.9% as compared to other machine learning methods, for instance, Support Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), Multilayer Perceptron (MLP), and Decision Tree (DT). Scientific analysis techniques (i.e., Fuzzy Delphi Method (FDM)) and machine learning feature engineering techniques (i.e., Variance Threshold (VT)) were used in two different experiments for selecting features where scientific analysis techniques had achieved better accuracy. The finding of this research is that, along with the past performance and social status key factors, the semester behavior factors have a lot of impact on students' performance. We performed spatial statistical analysis on our dataset in the context of Pakistan, which provided us with the spatial areas of students' performance; furthermore, their results are described in the data analysis section.

**Keywords:** student performance; Educational Data Mining; higher education; spatial analysis; predictive modeling; Fuzzy Delphi Method

# 1. Introduction

Quality education provides the basis for equality in society. One of the most basic public services is high-quality education [1]. Quality of education is vital for every citizen [2]. For this, educational institutes should be focused to improve the academic performance of every student individually. To achieve overall academic success, students need to perform well in all courses [3]. It is quite difficult for educators to keep track of their students' academic performance and improve their performance in each course [4]. As they cannot manage individual course-wise records manually, so they are unable to improve students' performance when it is required to meet the demands of each student on different attitudes [5]. Thus, a technical automated system is required which should provide detailed information on students' progression, which should input exam results, assignments, and



Citation: 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. https:// doi/10.3390/app121910112

Academic Editor: Amelia Zafra

Received: 31 July 2022 Accepted: 29 September 2022 Published: 8 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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/). performance of other activities in the course [6]. Researchers are working on a variety of statistical and machine learning models for predicting the impact of academic factors on different students [7].

Educational Data Mining (EDM) is a field that has been used to analyze academic data [8,9]. EDM has various applications during the development of a system as it uses different computational methods to detect patterns to analyze large-scale data. The prediction of student results in terms of academic performance, ratings, or grades is a well-known application of Educational Data Mining [10–13]. Predictive modeling techniques have been considered for students' academic performance [14,15], in this regard, classification techniques come out to be the most effective for this problem.

Data mining can be used in different fields to improve overall efficiency by using pattern analysis [16]. It is possible only by extraction of valuable information [17] from a stored dataset that is undiscovered until now. This extracted useful information will be used later, as it will help in resolving the issues that were previously faced during the development of the structural model. Today, data mining and its applications in the education sector have gained importance more than before. Thus, we can define Educational Data Mining as 'the process to transform raw data from the educational system into useful information that can be used later by the stakeholders for further applications' [18]. In the end, it will assist the educational institutes to review, improve, and strengthen students' learning process. It is obvious that to enhance the environment of any educational institute, the most important thing is to understand the learning process of students. The sheer understanding of this process has several advantages like optimization of learning outcomes for students and making the system strong enough to support weak students [19]. As a result, the rate of students failing their courses and dropping out will be decreased [20].

In the literature on geography education, spatial thinking is more closely tied to spatial skills, aptitudes, and ideas [21,22]. Choosing the best route to commute to work or school is only one example of how spatial thinking is used on a daily basis. Geographical Information systems (GIS), in particular, can improve spatial thinking because they make it possible to analyze geospatial data and find hidden patterns inside the data [23,24]. GIS is a system that is used for the management, storage, and analysis of geospatial data. GIS-based applications are mostly available online easily. These applications can be used by anyone for processing data concerning its spatial features [25].

Geospatial data is comprised of both location and characteristics of spatial features. To define a lane, for example, a reference is made to its position (i.e., where it is) and its attributes (e.g., length, name, speed limit, and direction). A GIS enables the user to handle road data and many other geospatial data, thereby separating them from non-spatial data business management systems. In addition to geospatial data, a GIS contains hardware, software, people, and organization. GIS hardware includes computers, printers, plotters, digitizers, scanners, Global Positioning System (GPS), and mobile devices. GIS software, either commercial or open-source, contains programs and applications for data management, data interpretation, data display, and other tasks to be performed by a computer [26].

Student dropout in the educational sector is a very important issue in higher education and If students' dropout rate is high then it will surely waste the resources of the institution and will also affect its credibility whenever an institutional evaluation will be performed [27]. Consequently, it is the need of the hour to propose a model that will output the estimation of the final result of the students by making use of their previous records to reduce the rate of dropout. This will also enhance the quality of education. For this, all the faculty members, administration, and educational system of the institutions should take this responsibility to design better outlines of learning and establish useful systems which will enhance learning opportunities for the students [28].

Hence, in this paper, firstly, we identified the student performance risk factors and semester behavioral factors from the literature in order to predict their performance.

Secondly, we conducted three experiments to meet the objectives. In the first experiment, we defined the student performance prediction by using a scientific analysis technique, which is the Fuzzy Delphi Method (FDM) for screening and shortlisting the student performance key factors. In the second experiment, we incorporated all identified risk factors for predicting student academic performance. Finally, in the third experiment, we used the machine learning feature engineering technique, which is the Variance Threshold (VT) for predicting student academic performance. The main objective of this paper is to first find and use the spatial locational factors and semester behavioral factors for predicting student academic performance. The key factors that have the most impact on student performance. The last objective is to analyze the spatial data in terms of spatial statistical analysis.

# 2. Related Work

The emergence of Education Data Mining (EDM), the latest discipline which has been used for over a decade in the development, study, and application of computerized methods in pattern detection, has helped exponentially in the analysis of vast educational data [29] that would otherwise be difficult due to large volume of existing data. The prediction of student results, where the aim is to evaluate the untold gain of function, information, ratings, or grades, is considered one of the experienced and famous applications of educational data mining (EDM) [30]. One of the historical student data findings, a predictive model for the success of student performance, is a highly recommended technique to investigate the relevant problems of students [31,32].

To increase the overall efficiency of a system, data mining (DM) can be applied in different fields. This can be achieved by extracting valuable and specific knowledge previously undiscovered from a stored data set [33]. In this way, the information learned will help solve several challenges and develop the current structure [34].

The use of DM in education is of increasing importance. In fact, for college learners, conventional DM techniques can be applied to educational data for the results. EDM is defined as the process [18] used to transform raw data collected by education systems into useful information that teachers can use to take corrective action and answer research questions. Thus, EDM assists education centers to review and strengthen students' learning processes. In enhancing an institution's educational environment, understanding students' knowledge-based learning should have played a huge part in developing skills. Such awareness results in many advantages, such as optimizing learning outcomes for students and the ability to prepare outcomes for the support of weak students. The number of students dropping out or failing classes would decreased as a consequence [35].

Estimating student performance is not an easy task and it is also important for both students and teachers to be aware of student performance. Early estimation is helpful for students and teachers. Teachers can play an important role for students and keep them aware of students dropping out of their course or subject in university. Teachers can also help students who need extra support [35].

Student dropout rates of academic students, which is one of the significant problems in higher education, affects the resources of the university and eventually affects the institutional evaluation process [27]. It is necessary to propose an evaluation model for the estimation of results for academic students. This will give support to the academic quality process and reduce dropout rates. We should give priority to education and communication in our societies. It is the responsibility of teachers and all education systems and their administrators to develop better outlines of learning and establish systems to expand learning opportunities [28].

We need to identify weak students among the whole class through their performance predictions using different techniques to provide the proper attention and to prevent them from dropping out of their studies [36]. Therefore, to support students' dropout rate, early warning systems need to be made and considered [37]. Due to incomplete and faulty

information systems of educational organizations, student behavioral characteristics are used for student performance predictions [38].

Locational features also have some impact on students' performance. In this regard, the geographic location of public schools has been considered [39], and the study concluded that the geographic location of public schools does affect the performance of academic students. Another study concluded that the rural graphical location of the teaching side was associated with high students satisfaction [40]. Yet another study on location highlighted their findings by concluding that school resources vary across graphical locations, and communities having small rural areas have the lowest socioeconomic profile [41], lower student academic performance, shortage of education staff, and industrial material, while schools in the neighborhood of towns have high resources, more availability of teaching staff, and higher students' satisfaction [42]. Another study [43] concluded that students are not performing well in remote areas as compared to their country counterparts. A study [44] further concluded that students who study in city location schools achieve significantly greater marks as compared to other geographic areas.

Student dropout rates also cause financial loss for both students and their education sectors. It affects graduation rates and lowers employment opportunities in highly qualified positions. If an institution loses a student, it decreases the retention rate of the university. Education for Sustainable Development is an important factor for making societies better, higher education guarantees any society to produce future professionals and leaders [28]. That is why the anticipation of good performance on the part of students is a significant study area as it can make students aware of their expected results before final exams.

This prediction will be an alert for weaker students that they have to put in extra effort than before, in order to achieve better results than predicted. If we apply this theory based on an institutional perspective, observe how, by performing different prediction techniques, these affected students will be identified, and as a result, their teachers can provide their full attention to ease their studies and keep them safe from dropping out. With the help of these predictions, we can make early warning systems to decrease student dropout rates [37].

# 3. Data Collection

#### 3.1. Student Performance Factors

Comprehensive evidence through literature was gathered. Researchers have been using numerous machine learning techniques for predicting the performance of academic students. In this regard, 33 studies have been identified from 2015 to 2022. The aim of collecting these studies was to find the key factors that have mostly been used or have been important to be used for predicting the academic performance of students.

Primarily, the comprehensive literature sorted out the student risk factors that have been identified by research articles for the prediction of student academic performance. Secondly, this literature was focused to include the key factors that have been highlighted by multiple educational articles, surveys, and systemic literature review articles. Finally, the literature was reviewed to find factors that have not been used in research articles for students' performance prediction. These features have been mentioned in the "New Features" column in the Taxonomy of Systematic Literature Review in Figure 1.



Figure 1. Taxonomy of Systematic Literature Review.

## 3.2. Students' Semester Behavioral Feature

Multiple students' semester behavioral features have been considered in this research. The list of these factors has been listed in Student performance factors and FDM output Table 1. A Google Form was designed considering the 24-semester behavioral features of students. At the end of their semester, students were asked to fill out this form. A total of 200 students took part in this survey, who have been targeted for their first four semesters. Later, a survey was conducted by experts to find the importance of these students' academic features. Such a process was monitored using the Fuzzy Delphi Method [45]. The features that have lower importance according to experts were discarded.

Table 1. Student performance factors and FDM output.

Input Feature	#	Description	FDM Score
Age	1	Current Age, ranges from 18–27 (10 Values)	3.06
Gender	1	Male or Female (Binary)	2.78
Family Income	1	(Multiple Values)	3.52
Guardian Education	1	Education of Guardian (5 values)	3.41
Behavior of Society/Surrounding Environment	1	Rank of Society (5 values)	3.84
District of Origin	1	Districts (Multiple values)	2.75

Input Feature	#	Description	FDM Score
Geographic Location	1	Urban or Rural (Binary)	3.52
Access to Education	1	Distance between home and university (Multiple Values)	3.02
Quality of Education	2	Quality of education in school and college (5 values)	3.75
Type of Institute	2	Type of school and college, Good or Bad (Binary)	4.05
Financial Aid	1	Financial Aid or Scholarship, Yes or No (Binary)	3.63
Teaching Quality	1	Teaching quality (5 Values)	3.98
Attendance	1	Attendance Percentage (Multiple Values)	3.95
Geography of Educational Institute	2	Geography of school and college, Urban or Rural (Binary)	3.46
Past Academic Performance	2	10th Grade Average, 12th Grade Average (Multiple Values)	2.93
Hostellers/Day Scholars	1	Yes or No (Binary)—Semester Wise	3.19
Self-Employed	1	Yes or No (Binary)—Semester Wise	3.55
Married/Un-married	1	Yes or No (Binary)—Semester Wise	3.48
Way of Coming to university	1	Financial Aid or Scholarship, Yes or No (Binary)	2.89
Personal Issues	1	Yes or No (Binary)—Semester Wise	3.82
Home Issues	1	Yes or No (Binary)—Semester Wise	3.76
Group-wise study	1	Yes or No (Binary)—Semester Wise	3.14
Make new friends	1	Yes or No (Binary)—Semester Wise	2.91
Help from seniors	1	Yes or No (Binary)—Semester Wise	3.18
Hardworking friends	1	Yes or No (Binary)—Semester Wise	3.52
Connection with religion	1	Yes or No (Binary)—Semester Wise	3.05
Elder sibling study same course	1	Yes or No (Binary)—Semester Wise	3.63
Elder sibling still studying	1	Yes or No (Binary)—Semester Wise	3.59
Spend time on social media	1	Yes or No (Binary)—Semester Wise	3.13
Spend time in gym/exercise	1	Yes or No (Binary)—Semester Wise	3.09
Study daily basis	1	Yes or No (Binary)—Semester Wise	3.80
Study weekly basis	1	Yes or No (Binary)—Semester Wise	3.50
Study for quizzes	1	Yes or No (Binary)—Semester Wise	3.65
Make assignments	1	Yes or No (Binary)-Semester Wise	3.68
Bunk lectures	1	Yes or No (Binary)—Semester Wise	2.99
Bunk university	1	Yes or No (Binary)-Semester Wise	3.00
Internet quality at home	1	Yes or No (Binary)—Semester Wise	3.38

# Table 1. Cont.

## 3.3. Fuzzy Delphi Method

A systematic review of the student performance factors revealed that there are many key factors. As it is impractical to consider all input factors for the student performance prediction process, a proper screening approach must be utilized to identify the significant factors between a larger set of inputs. Delphi is an extensively used method for screening that tries to seek the most critical or influential elements of a phenomenon under consideration. However, there are certain demerits of the traditional Delphi method including less uniformity of fine judgments, increased computational cost, and up-gradation of skilled personal judgments to attain uniform altogether judgments [46]. The Fuzzy Delphi Method [47] addresses these demerits and has received significant popularity since its inception. This research utilizes the Fuzzy Delphi Method to screen out the most significant input factors for student academic performance prediction.

The Fuzzy Delphi Method (FDM) was employed in the screening and shortlisting of input features for student performance prediction. The threshold value was determined through this process to seek expert judgments regarding the significance of the evaluation factors in the list [48]. A questionnaire survey is utilized as the main research tool for the data collection in this phase. This questionnaire comprises two sections, where the first section collected the basic information and demographics of the experts. The second section collected the opinions of the experts regarding the significance of a particular input factor for student performance prediction. Considering a large number of factors, the second part of the questionnaire is divided into five sub-sections guided by the underlying perspectives. A Likert scale of 1–5 is used to evaluate the factors where the higher point refers to the

higher significance of a factor. The research population comprised educationists from the educational domain. Finally, considering the guidelines and requirements of Fuzzy Delphi screening, questionnaires were distributed through Google Forms. A comprehensive list of 44 factors concerning student performance was formulated depending on results acquired from existing literature, technical papers, and blogs. Later, skillful responses for these factors were collected from 42 educational professionals.

The fuzzy Delphi screening or evaluation phase consists of three constituent steps including: (i) conversion of judgments into estimates, (ii) defuzzification using Graded Mean Integration Representation (GMIR), and (iii) screening of the critical factors according to the threshold value. The threshold can be subjectively set according to the mean or geometric mean of all evaluation factors [45]. By referring to the geometric mean of all 44 candidate evaluation factors, the threshold value was subjectively set to 3.37. Based on results from the Fuzzy Delphi screening process, a detailed and multidisciplinary grading system was developed. Out of 44 features, 29 features crossed the threshold mark of the grading system which was 3.37 out of 5, and were considered significant.

Certain unavailable features such as Family Income and Attendance rendered the dataset with 27 features. Features such as teaching quality was thought to be entered from only our university where it was considered to be the same for each student. The Fuzzy Delphi thresholding discarded this feature as well. As a result, the dataset was left with 26 features.

Table 1 presents all input features that are incorporated in this research with their number of meanings, their feature description, and their FDM Score (feature threshold value). Thirty-seven features are listed in this table. Out of these 37 features, four features are used twice, because of two meanings, mentioned in the column (#) and column (description) of the same Table 1. There are three features from the list of educational factors of Figure 1 that were dropped in the data preprocessing phase because they also existed in the behavioral factors list of Figure 1. These three features are Self Employed, Marital Status, and Hostel Factor.

### 3.4. Dataset

In this research, we used the dataset of the FAST National University of Computer and Emerging Sciences (NUCES) Chiniot-Faisalabad (CFD) campus. The dataset of 200 students in their first four semesters of bachelor's degree in the discipline of Computer Science is considered for this research. The data features included academic features, locational features, and semester behavioral features, listed in Table 1. The dataset consists of 47 input features and 800 records. Three new features that were not incorporated by the Fuzzy Delphi process but are included in this dataset are Guardian (Father, Mother, or Other), attempted credit hours (Multiple Values), and earned credit hours (Multiple Values). In the data cleaning process, due to some missing information from a record, we had to remove a record. So dataset was left with 799 records for this research.

#### 3.5. Data Analysis

The histogram in Figure 2 shows the semester-wise behavior features of students. From the histogram, we can estimate the score varies between semesters. Some of the factors that have the higher score in FDM have been highlighted in Figures 2 and 3. Behavior features include the new students' urge to make new friends in their first semester, hosteller trends in their first four semesters, personal and home issues faced by the students, students having older siblings in the same field, and some students with other factors. The results showed that over 80%, of students make new friends in their first semester, and here we have extracted fruitful information from students' semester behavioral records.



Figure 2. Semester-wise Behavioral features of Students (1).

A comparison of graphs between daily study or weekly study habit of students is also made. Older siblings in education, internet facility at home, and guidance from seniors are shown in the histogram. The results show that the semester 4 graph has high values during the period of COVID-19.





The overall behavioral data of students has also been analyzed in Different Input Parameters, as shown in Figure 4. Some important parameters have been visualized here for a better understanding of the students' behavioral data. The percentage of male students in this dataset is higher compared to female students. A very low number of student are self-employed during their studies. Different types of students came from different backgrounds/locations. Some other parameters that are visualized in Figure 4 are SSC (secondary school certificate equivalent to the 10th standard), HSSC (higher secondary school certificate equivalent to the 12th standard), O-Level (qualification in a specific subject equivalent to the 10th standard), and A-Level (a qualification that is equal to the 12th standard).

Geo-mapping was then performed by using the coordinates data of students. Geo-mapping is the act of translating locational coordinates into a geo map that is used to visualize the location of utilities quickly and correctly. It is a method for displaying data from various geo-cultural contexts or specific geographic areas. It is also used to make maps of real-world features' spatial locations and to illustrate the spatial relationships between them.



Figure 4. Different Input Parameters.

For this study, the student data was visualized through geo-mapping as shown in Figure 5 (Multivariate clustering of students data). Figure 5 shows students' locality within different regions of Pakistan. The zoomed map of the state, Punjab of Pakistan, also indicated that we have a majority of students from metropolitan areas like Faisalabad and Lahore. Multivariate clustering of the Punjab boundary shows the clusters based on the performance of students in the region of Punjab.



Figure 5. Multivariate clustering of students data (Pakistan and Punjab Boundary).

#### 4. Proposed Methodology and Results

The proposed methodology includes three experiments for the prediction of student performance. The first experiment used the Fuzzy Delphi Method output for the prediction of student performance. The second experiment is applied to all the datasets having all academic features, locational features, and behavioral features. The last experiment applies different feature selection techniques to the complete dataset followed by a trial that tried to predict student performance.

#### 4.1. Methodology

In the methodology phase, multiple steps are encountered for data processing. In the data preprocessing phase, we performed different steps, starting with data cleaning. In the data cleaning step, a record was removed because we did not have their semester information. A record contains the students' data for a semester, and because of the unavailability of some important features, we had to remove that record before using the data in the experiments. Data binning was performed on some columns containing numeric values, such as the feature 'Past Performance'. There are some categorical columns, having values 'Yes' and 'No', directly normalized to 1 and 0 in the data normalization. In the last phase of data preprocessing, label encoding is performed on such columns, where a count of unique column value is more than two. Our dataset has a multi-label class (Good-Performer, Avg-Performer, Bad-Performer). To synthesize the dataset, SMOTE (Synthetic Minority Oversampling Technique) is used. Before applying SMOTE to the dataset, the dataset had 799 records and 47 features, and after applying SMOTE, dataset records increased to 1428 with the same 47 features (Figure 6).



Figure 6. Dataset Records without-SMOTE and with-SMOTE.

We performed three experiments to predict the academic performance of students (Figure 7); for all experiments we performed the same preprocessing steps (Data bining, Label encoding, Data normalization, and Data Synthesis using SMOTE). The aim of the first experiment (Exp-1) is to consider only those features that have higher importance according to experts and to obtain their consensus on the sustainability of the presented item in the questionnaire. In this experiment, we have used the Fuzzy Delphi method (FDM) which is a scientific analysis technique to consolidate consensus agreement within the panel of experts. FDM was used to shortlist these 47 features. After taking consensus from 42 experts, we were left with 26 features for experiment one (Exp-1). Later, after performing the preprocessing steps (Data bining, Label encoding, Data normalization, and Data Synthesis using SMOTE), we had two datasets for experiment one (Exp-1). For without-SMOTE, the dataset had 799 records and 66 features, and for with-SMOTE, the dataset had 1428 records and 66 features.





Figure 7. Dataset Properties.

For the second experiment (Exp-2), the aim is to consider all the features that have been called important by the literature. In this research, we collected all 47 important features, and we used all these features in this experiment. After performing different preprocessing, we obtained two datasets for experiment two (Exp-2) as well. For without-SMOTE, the dataset had 799 records and 116 features, and for with-SMOTE, the dataset had 1428 records and 116 features.

Lastly, the third experiment (Exp-3) aims to consider the features that have higher importance according to machine learning features selection techniques. In this experiment, we applied three machine learning feature engineering techniques. Feature engineering techniques Select K-Best, Variance Threshold, and L1 Based techniques are applied in this experiment. After obtaining the results from these three-feature engineering techniques in the pilot experiment, we analyzed that the variance threshold feature engineering techniques. The accuracies achieved with variance threshold are much better than the other two techniques, as can be seen from Figure 8.



Figure 8. Exp3 Pilot Experiment.

As a result, we used the method of Variance threshold. Here again, we have two datasets in the third experiment (Exp-3) after performing different preprocessing steps. For without-SMOTE, the dataset has 799 records and 43 features, and the with-SMOTE dataset has 1428 records and 37 features. Here, features vary in both methods (without-SMOTE and with-SMOTE) because the variance threshold feature selection technique is applied separately to both data sources (without-SMOTE data and with-SMOTE data). Figure 9 elaborates the dataset properties of all the experiments.

In this research, we used different machine learning and deep learning algorithms for the prediction of academic student performance. These algorithms are Naïve Bayes (NB), Decision Tree (DT), Long Short-Term Memory (LSTM), Multi-Layer Perceptron (MLP), Random Forest (RF), and Support Vector Machine (SVM). Hyperparameter tuning in the shape of Grid Search CV is also applied in all these experiments, using both without-SMOTE and with-SMOTE. Different data evaluation techniques have been applied to evaluate the performance of the models. These evaluation techniques are Accuracy, Precision, Recall, F1-Score, and ROC (Receiver Operating Characteristic) curve.

Spatial statistical analysis techniques have also been applied to find the spatial behavior of the dataset. We used two methods: Multivariate Clustering and Average Nearest Neighbor.



Figure 9. Proposed Methodology.

#### 4.2. Results

We carried out three experiments to test our methodology and to compare the performance of different proposed experiments with state-of-the-art machine learning and deep learning classification algorithms.

#### 4.2.1. Experiment 1 (Exp-1)

Experiment 1 is performed on Fuzzy Delphi output. The input features that are shortlisted by applying FDM are used for student academic performance prediction. In this experiment, the Support vector machine (SVM) achieved the best accuracy compared to Decision tree (DT), Long short-term memory (LSTM), Multi-layer perceptron (MLP), Naïve Bayes (NB), and Random Forest (RF). This experiment is performed without SMOTE and with SMOTE. SMOTE is used for data balancing. With SMOTE, we can see that SVM obtains a higher accuracy of 89.5 as compared to all other models especially Random Forest and Long short-term memory, as shown in Figure 10.



Figure 10. Exp1 Models Comparison.

Regarding precision, recall, and F1-score, here again, the support vector machine (SVM) performs well with SMOTE; the multi-layer perceptron obtains good results with SMOTE as compared with all models especially Random Forest (RF) and deep learning model long short term memory (LSTM) (Table 2).

Exp 1	Without-SMOTE		With-SMOTE		Έ	
Algorithms	Precision	Recall	F1-Score	Precision	Recall	F1-Score
DT	0.57	0.58	0.57	0.78	0.79	0.78
LSTM	0.7	0.63	0.66	0.86	0.86	0.86
MLP	0.73	0.67	0.69	0.86	0.86	0.86
NB	0.51	0.6	0.44	0.73	0.65	0.62
RF	0.76	0.6	0.65	0.88	0.87	0.87
SVM	0.71	0.65	0.67	0.9	0.9	0.89

Table 2. Exp1 Scores (Precision, Recall, F1-Score).

After applying hyperparameter tuning by using grid search CV, we also applied hyperparameter tuning by considering both data sources (without-SMOTE data, and with-SMOTE data). We retrieved the best hyperparameters for each used machine learning algorithm of both data sources (Table 3).

Table 3. Exp1 Scores (Precision, Recall, F1-Score).

Algorithms	Without-SMOTE	With-SMOTE
	Criterion = gini,	Criterion = entropy,
ЪΤ	<pre>min_samples_leaf = 1,</pre>	min_samples_leaf = 1,
DI	<pre>min_samples_split = 2,</pre>	<pre>min_samples_split = 2,</pre>
	splitter = best	splitter = best
	Activation = relu, batch_size = 7,	Activation = relu, batch_size = 17,
	epochs = 120, last_act = softmax,	epochs = 120, last_act = softmax,
LSTM	solver = adam, unit-LSTM = 95,	solver = Nadam, unit-LSTM = 80,
	unit-Dense = $95$ , dropout = $0.2$	unit-Dense = 45 dropout = 0.1
	[(1+1) LSTM, Dense]	[(4+1) LSTM, Dense]
	Activation = $tanh$ , $alpha = 0.001$ ,	Activation = relu, $alpha = 0.1$ ,
MLP	hidden_layer_sizes = (95, 95, 95,	hidden_layer_sizes = (87),
	95), max_iter = 200, solver = adam	max_iter = 200, solver = adam
NB	Default	Default
	Bootstrap = False,	Bootstrap = False,
DE	$min_samples_leaf = 1,$	$min_samples_leaf = 1,$
КГ	min_samples_split = 5,	min_samples_split = 2,
	$n_{estimators} = 50$	$n_{estimators} = 500$
CVDA	C = 100, Gamma = 0.01,	C = 10 Commo = 0.5 kornol = rhf
<i>3</i> v IVI	kernel = rbf	C = 10, Gamma = 0.3, Kerner = 101

## 4.2.2. Experiment 2 (Exp-2)

Experiment 2 is performed on the full dataset, which has 799 records and 47 features. In this experiment, Random Forest (RF) obtained higher accuracy using without-SMOTE as compared to Decision tree (DT), Long short-term memory (LSTM), Multi-layer perceptron (MLP), Naïve Bayes (NB), and Support vector machine (SVM). This experiment is performed without SMOTE and with SMOTE. With SMOTE, we can see that LSTM achieved a higher accuracy of 90.9 as compared to other models, as shown in Figure 11.



Figure 11. Exp2 Models Comparison.

With regard to precision, recall, and f1-score, LSTM performed better as compared to other models in both data sources (without-SMOTE data, with-SMOTE data) (Table 4).

Exp 2	Wi	Without-SMOTE		W	ith-SMOT	Έ
Algorithms	Precision	Recall	F1-Score	Precision	Recall	F1-Score
DT	0.68	0.69	0.69	0.86	0.86	0.86
LSTM	0.77	0.74	0.73	0.91	0.91	0.91
MLP	0.75	0.72	0.73	0.89	0.89	0.89
NB	0.53	0.62	0.39	0.68	0.58	0.54
RF	0.78	0.65	0.67	0.89	0.89	0.89
SVM	0.74	0.71	0.72	0.88	0.88	0.88

Table 4. Exp2 Scores (Precision, Recall, F1-Score).

Detailed hyperparameter tuning of LSTM was performed in this experiment. Firstly, the number of neurons of a single LSTM layer was identified with an epoch value of 70 and batch-size value of 10. Then, the number of LSTM layers was identified by providing the best number of neurons with the same epochs and batch-size values. In the next phase, we added and found the best number of dense layers (fully connected layers) in the LSTM model, with the best-identified LSTM layers and neurons, and with the same epochs and batch-size values. In the fourth phase, the best number of neurons in the dense layer(s) were identified. In the fifth and sixth phases, several epochs and batch sizes were identified by providing them with the best values of LSTM layers, LSTM neurons, Dense layers, and Dense neurons. This experiment was performed with both data sources (without-SMOTE data and with-SMOTE data), but we are visualizing the working of the with-SMOTE data, as it achieved the highest accuracy (Figure 12).

LSTM loss and accuracy measure were visualized with both the test and train values of the dataset, and ROC curve were plotted to evaluate the LSTM model (Figure 13). The ROC Curve shows good results with regard to the LSTM model. The hyperparameter of this experiment is mentioned in Table 5.



Figure 12. Exp2 LSTM Hyper Parameter Tuning.



Figure 13. Exp2 (LSTM Accuracy and Loss) and ROC Curve.

Algorithms	Without-SMOTE	With-SMOTE
	Criterion = entropy,	Criterion = entropy,
Т	<pre>min_samples_leaf = 1,</pre>	<pre>min_samples_leaf = 1,</pre>
DI	<pre>min_samples_split = 5,</pre>	<pre>min_samples_split = 5,</pre>
	splitter = best	splitter = best
	Activation = relu, batch_size = 5,	Activation = relu, batch_size = 17,
	epochs = 50, last_act = softmax,	epochs = 120, last_act = softmax,
LSTM	solver = adam, $unit-LSTM = 95$ ,	solver = Nadam, unit-LSTM = $85$ ,
	unit-Dense = $60$ , dropout = $0.1$	unit-Dense = $80$ , dropout = $0.1$
	[(1+2) LSTM, Dense]	[(1+2) LSTM, Dense]
	Activation = $tanh$ , $alpha = 0.1$ ,	Activation = $tanh$ , $alpha = 0.1$ ,
MLP	hidden_layer_sizes = (93,93),	hidden_layer_sizes = (77,77),
	max_iter = 200, solver = adam	max_iter = 100, solver = adam
NB	Default	Default
	<pre>min_samples_leaf = 2,</pre>	min_samples_leaf = 2,
RF	<pre>min_samples_split = 3,</pre>	<pre>min_samples_split = 3,</pre>
	$n_{estimators} = 100$	$n_{estimators} = 100$
SVM	C = 100, Gamma = 0.01,	C = 100, Gamma = 0.01,
5 1 11	kernel = rbf	kernel = rbf

Table 5. Exp2 Scores (Precision, Recall, F1-Score).

### 4.2.3. Experiment 3 (Exp-3)

Experiment 3 is performed on the full dataset, which has 799 records and 47 features in the beginning. In this experiment, different machine learning feature selection techniques have been applied. Variance Threshold provided the best performance in the pilot experiment with all machine learning models as compared to other feature engineering techniques (Select K Best and L1-based feature engineering). For this, in experiment 3, we have considered this feature selection technique.

In this experiment, Support vector machine (SVM) achieved higher accuracy using without-SMOTE as compared to Decision tree (DT), Long short-term memory (LSTM), Multi-layer perceptron (MLP), Naïve Bayes (NB), and Random Forest (RF). This experiment was performed without SMOTE and with SMOTE. With SMOTE, we can see that RF achieved a higher accuracy of 88.8 as compared with other models (Figure 14).



Figure 14. Exp3 Models Comparison.

With regard to precision, recall, and f1-score, here again, the Random Forest achieved the highest accuracy with SMOTE, as shown in Table 6.

Exp 3	Without-SMOTE		With-SMOTE		Έ	
Algorithms	Precision	Recall	F1-Score	Precision	Recall	F1-Score
DT	0.65	0.61	0.62	0.83	0.83	0.83
LSTM	0.69	0.68	0.68	0.88	0.87	0.87
MLP	0.65	0.64	0.65	0.86	0.86	0.86
NB	0.52	0.48	0.49	0.69	0.69	0.69
RF	0.76	0.58	0.59	0.89	0.88	0.88
SVM	0.72	0.63	0.64	0.88	0.88	0.88

Table 6. Exp2 Scores (Precision, Recall, F1-Score).

After applying hyperparameter tuning by using grid search CV, we retrieved the best hyperparameters for each model, as shown in Table 7. Because we applied hyperparameter tuning on both data sources (without-SMOTE data, and with-SMOTE data), we obtained the best hyperparameters for each model for both data sources (Table 7).

Table 7. Exp2 Scores (Precision, Recall, F1-Score).

Algorithms	Without-SMOTE	With-SMOTE
DT	Criterion = gini, min_samples_leaf = 2, min_samples_split = 5, splitter = best	Criterion = gini, min_samples_leaf = 1, min_samples_split = 2 splitter = best
LSTM	Activation = relu, batch_size = 7, epochs = 110, last_act = softmax, solver = Nadam, unit-LSTM = 30, dropout = 0.2 [(1+0) LSTM, Dense]	Activation = relu, batch_size = 5, epochs = 100, last_act = softmax, solver = adam, unit-LSTM = 100, unit-Dense = 65, dropout = 0.2 [(1+3) LSTM, Dense]
MLP	Activation = tanh, alpha = 0.001, hidden_layer_sizes = (93,93,93), max_iter = 100, solver = adam	AActivation = tanh, alpha = 0.001, hidden_layer_sizes = (93,93,93,93), max_iter = 100, solver = adam
NB	Default	Default
RF	<pre>min_samples_leaf = 2, min_samples_split = 2, n_estimators = 200</pre>	<pre>min_samples_leaf = 2, min_samples_split = 2, n_estimators = 120</pre>
SVM	C = 2, Gamma = 0.1, kernel = rbf	C = 2, Gamma = 0.1, kernel = rbf

# 4.3. Results Comparison

We have performed three different experiments on our research problem and accuracies of the experiments have been compared. Using Without-SMOTE data, we achieved the highest accuracy with experiment 2 by Random Forest (RF); using with-SMOTE data, the highest accuracy was also achieved with experiment 2 by Long Short Term Memory (LSTM), as shown in Figure 15 and the compared results in Table 8.

Table 8. Experiments Comparison (Accuracies).

	Experir	nent 1	Experin	nent 2	Experin	nent 3
Algorithms	Without-SMOTE	With-SMOTE	Without-SMOTE	With-SMOTE	Without-SMOTE	With-SMOTE
DT	63.7	79	75	86	72.5	83.2
LSTM	74.4	86.4	78.8	90.9	75	87.4
MLP	75.6	86	78.8	88.8	72.5	86
NB	41.9	66.8	38.1	60.1	58.1	69.2
RF	75	87.1	80	89.5	77.5	88.8
SVM	74.4	89.5	78.1	88.1	78.1	88.5



Figure 15. Experiments Comparison.

# 4.4. Significance of Features (P-Value)

In our research context, eleven features have a significant p-value (value < 0.05). Features like past performance, society status, and semester behavior have a lot of impact on students' performance (Figure 16).





#### 5. Conclusions and Future Works

In the current study, our main concern was to predict students' academic performance at an early stage of the semester so that early predictions will make students aware of their expected results, and early warning systems can be made to support student dropout rates. This study also performed extensive literature and tried to find the importance of key factors that can play important role in the academic student's performance predictions. Moreover, this study has focused on the locational factors of the students that can play important role in the student's academic life. By the locational features, we can find out the areas or regions that are lacking or that need an uplift, so that proper educational facilities can be provided to them.

Our study was focused on finding and working on important key features, from which we can predict student performance at early stages. We tried to collect all key factors that had been highlighted by relevant articles. Later, we took consensus from the educational experts to give each factor a score by using a 5-point Likert scale. To obtain the threshold of the scores, we applied Fuzzy Delphi Method to the scores. We used all the factors whose scores exceeded the threshold value. The purpose of this process was to highlight all important factors that have effects on student performance. Hence, future researchers do not need to find the importance of educational factors again. Educational Institutes also work on these factors and predict early student performance so that they too can minimize the dropout rates.

Another main aspect of our study was to consider locational factors, as it was one of our goals to find out the importance of some locational features (e.g., student location (urban, rural), access to school distance, society status, and geo coordinates of student location). We also performed some GIS analytics to discover the areas or regions that are lacking or that need an uplift, so that proper educational facilities can be provided to them.

Our study predicted student academic performance with high accuracy at early stages and highlighted key factors that are affecting student performance. Thus, these features are important not only to predict the academic performance of students but also to decrease dropout rates, increase graduation rates, drop the financial loss of both students and educational sectors, and most importantly provide high employment rates.

In our findings, the deep learning model LSTM achieved the highest accuracy of 90.9 compared with state-of-the-art machine learning algorithms. LSTM performed better when features and dataset records were large in number. Student academic performance prediction is performed with three experiments: (a) using the Fuzzy Delphi Method; (b) using educational key factors; and (c) using Machine Learning feature engineering techniques. The SMOTE data synthesizing tool was considered a significant method to deal with the unbalanced nature of the dataset for student performance prediction in all the experiments. According to this research, we can conclude that all the factors considered in the study have a higher correlation. We will obtain maximum accuracy if we use all factors for the prediction of students' academic performance. In this specific research context, we can conclude that the scientific analysis technique FDM obtains better accuracy as compared to the machine learning feature engineering technique (Variance K Threshold). Along with past performance and social status, the semester behavior factors have much impact on students' performance (T-Test). Spatial statistical analysis provided us the spatial information about the results (Performance areas and spatial correlation of features with Average Nearest Neighbor).

Estimation of student performance demands the attention of both students and teachers towards student performance where teachers can play an integral part by helping those students who need extra support and keeping them aware of dropping out of their course or subject. Student dropout rate is a potential problem that causes financial loss to both students and education sectors, affects graduation rates, and lowers the employment opportunities in highly qualified positions. Thus, we proposed a model based on previous records of students that will help estimate the final result of students and help reduce the rate of student dropout.

Substantial work has been performed on the prediction of the performance of academic students using some data mining and machine learning models, where less importance was given to the locational features for prediction. In this research, we have combined both geospatial and machine learning tools for creating a relationship between students' location factors with their academic performance. The main purpose was to establish a GIS-based system that takes the geographic location of students, evaluates various educational strategies based on machine learning techniques, and generates results using multiple input data, for the prediction of their academic performance.

This study focused on finding the key factors, predicting the performance, and clustering the students in different areas based on their data class. Here, we especially used the geospatial locations of students. This study could not consider geo-socio-demographics features as we did not have the data. In the future, we can extend this work and predict student academic performance by considering their geo-locational attributes. As we have the coordinates data of students in our dataset, we can increase the amount of the dataset, and with the coordinates data of the student, we can extract their geo-locational features

or area-specific features (i.e., number of schools, universities, hospitals, etc.) to find the socio-demographics features of any region. This method of working on the geo-locations of the students will also help us to look at student performance prediction with a new and broader perspective.

**Author Contributions:** Conceptualization, M.H.M., R.M., N.S. and A.A.; Data curation, M.H.M.; Formal analysis, M.H.M.; Funding acquisition, M.S.S. and M.A.; Investigation, R.M. and N.S.; Methodology, M.H.M., M.S.S. and A.A.; Supervision, R.M. and M.A.; Validation, M.S.S., N.S. and A.A.; Visualization, N.S. and M.A.; Writing—original draft, M.S.S., R.M., N.S., A.A. and M.A.; Writing—review & editing, M.S.S., R.M. and M.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research study is funded by Higher Education Commission (HEC), Pakistan. The Project Title is "Vector-borne infectious Diseases Surveillance and Response System using Earth Observation Data" under the Project ID "NRPU-10537".

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

#### References

- 1. Lau, E.; Sun, L.; Yang, Q. Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Appl. Sci.* **2019**, *1*, 982. [CrossRef]
- 2. Saa, A.A.; Al-Emran, M.; Shaalan, K. Factors affecting students' performance in higher education: A systematic review of predictive data mining techniques. *Technol. Knowl. Learn.* **2019**, *24*, 567–598.
- Buenaño-Fernández, D.; Gil, D.; Luján-Mora, S. Application of machine learning in predicting performance for computer engineering students: A case study. *Sustainability* 2019, 11, 2833. [CrossRef]
- 4. Ma, Y.; Cui, C.; Yu, J.; Guo, J.; Yang, G.; Yin, Y. Multi-task MIML learning for pre-course student performance prediction. *Front. Comput. Sci.* **2020**, *14*, 145313. [CrossRef]
- 5. Hussain, M.; Zhu, W.; Zhang, W.; Abidi, S.M.R.; Ali, S. Using machine learning to predict student difficulties from learning session data. *Artif. Intell. Rev.* 2019, 52, 381–407. [CrossRef]
- 6. Wakelam, E.; Jefferies, A.; Davey, N.; Sun, Y. The potential for student performance prediction in small cohorts with minimal available attributes. *Br. J. Educ. Technol.* **2020**, *51*, 347–370. [CrossRef]
- Wang, P.; Li, L.; Wang, R.; Xie, Y.; Zhang, J. Complexity-based attentive interactive student performance prediction for personalized course study planning. *Educ. Inf. Technol.* 2022, 27, 5471–5493. [CrossRef]
- 8. Baker, R.S.; Yacef, K. The state of educational data mining in 2009: A review and future visions. J. Educ. Data Min. 2009, 1, 3–17.
- 9. Romero, C.; Ventura, S. Data mining in education. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2013**, *3*, 12–27. [CrossRef]
- 10. Romero, C.; Ventura, S. Educational data mining: A survey from 1995 to 2005. Expert Syst. Appl. 2007, 33, 135–146. [CrossRef]
- 11. Romero, C.; Ventura, S. Educational data mining: A review of the state of the art. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* 2010, 40, 601–618. [CrossRef]
- 12. Wolff, A.; Zdrahal, Z.; Herrmannova, D.; Knoth, P. Predicting student performance from combined data sources. In *Educational Data Mining*; Springer: Basel, Switzerland, 2014; pp. 175–202.
- 13. Yoo, J.; Kim, J. Can online discussion participation predict group project performance? Investigating the roles of linguistic features and participation patterns. *Int. J. Artif. Intell. Educ.* **2014**, *24*, 8–32. [CrossRef]
- Hämäläinen, W.; Vinni, M. Classifiers for educational data mining. In *Handbook of Educational Data Mining*; Chapman & Hall/CRC Data Mining and Knowledge Discovery Series; CRC Press: Boca Raton, FL, USA, 2011; pp. 57–71.
- Vialardi, C.; Chue, J.; Peche, J.P.; Alvarado, G.; Vinatea, B.; Estrella, J.; Ortigosa, Á. A data mining approach to guide students through the enrollment process based on academic performance. User Model. -User-Adapt. Interact. 2011, 21, 217–248. [CrossRef]
- 16. 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]
- 17. Shreem, S.S.; Turabieh, H.; Al Azwari, S.; Baothman, F. Enhanced binary genetic algorithm as a feature selection to predict student performance. *Soft Comput.* 2022, *26*, 1811–1823. [CrossRef]
- 18. Funatsu, K. New Fundamental Technologies in Data Mining; BoD–Books on Demand: Norderstedt, Germany, 2011.
- Alamri, R.; Alharbi, B. Explainable student performance prediction models: A systematic review. *IEEE Access* 2021, 9, 33132–33143. [CrossRef]
- Chung, J.Y.; Lee, S. Dropout early warning systems for high school students using machine learning. *Child. Youth Serv. Rev.* 2019, 96, 346–353. [CrossRef]
- Scholz, M.A.; Huynh, N.T.; Brysch, C.P.; Scholz, R.W. An evaluation of university world geography textbook questions for components of spatial thinking. J. Geogr. 2014, 113, 208–219. [CrossRef]
- 22. Jo, I.; Hong, J.E. Effect of learning GIS on spatial concept understanding. J. Geogr. 2020, 119, 87–97. [CrossRef]
- Gold, A.U.; Pendergast, P.M.; Ormand, C.J.; Budd, D.A.; Stempien, J.A.; Mueller, K.J.; Kravitz, K.A. Spatial skills in undergraduate students—Influence of gender, motivation, academic training, and childhood play. *Geosphere* 2018, 14, 668–683. [CrossRef]

- 24. Duarte, L.; Teodoro, A.C.; Gonçalves, H. Evaluation of Spatial Thinking Ability Based on Exposure to Geographical Information Systems (GIS) Concepts in the Context of Higher Education. *ISPRS Int. J. -Geo-Inf.* **2022**, *11*, 417. [CrossRef]
- Chang, K.T. Geographic information system. International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology. 2016, pp. 1–9. Available online: <a href="https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118786352.wbieg0152">https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118786352.wbieg0152</a> (accessed on 30 August 2020)
- 26. Folger, P. Geospatial Information and Geographic Information Systems (GIS): Current Issues and Future Challenges; DIANE Publishing: Darby, PA, USA, 2010.
- Paura, L.; Arhipova, I. Cause analysis of students' dropout rate in higher education study program. *Procedia-Soc. Behav. Sci.* 2014, 109, 1282–1286. [CrossRef]
- Mulà, I.; Tilbury, D.; Ryan, A.; Mader, M.; Dlouhá, J.; Mader, C.; Benayas, J.; Dlouhỳ, J.; Alba, D. Catalysing change in higher education for sustainable development: A review of professional development initiatives for university educators. *Int. J. Sustain. High. Educ.* 2017, 18, 798–820. [CrossRef]
- 29. Cortez, P.; Silva, A.M.G. Using data mining to predict secondary school student performance; EUROSIS-ETI, ETI Bvba: Ostend, Belgium 2008.
- Khan, A.; Ghosh, S.K. Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Educ. Inf. Technol.* 2021, 26, 205–240. [CrossRef]
- 31. Hoffait, A.S.; Schyns, M. Early detection of university students with potential difficulties. *Decis. Support Syst.* 2017, 101, 1–11. [CrossRef]
- Xiao, S.; Yan, J.; Farajtabar, M.; Song, L.; Yang, X.; Zha, H. Learning Time Series Associated Event Sequences With Recurrent Point Process Networks. *IEEE Trans. Neural Netw. Learn. Syst.* 2019, 30, 3124–3136. doi: [CrossRef]
- Li, L.; Yan, J.; Zhang, Y.; Zhang, J.; Bao, J.; Jin, Y.; Yang, X. Learning Generative RNN-ODE for Collaborative Time-Series and Event Sequence Forecasting. *IEEE Trans. Knowl. Data Eng.* 2022, 1–18. doi: [CrossRef]
- Miguéis, V.L.; Freitas, A.; Garcia, P.J.; Silva, A. Early segmentation of students according to their academic performance: A predictive modelling approach. *Decis. Support Syst.* 2018, 115, 36–51. [CrossRef]
- Lee, S.; Chung, J.Y. The machine learning-based dropout early warning system for improving the performance of dropout prediction. *Appl. Sci.* 2019, *9*, 3093. [CrossRef]
- 36. Skalka, J.; Drlik, M. Automated assessment and microlearning units as predictors of at-risk students and students' outcomes in the introductory programming courses. *Appl. Sci.* **2020**, *10*, 4566. [CrossRef]
- Márquez-Vera, C.; Cano, A.; Romero, C.; Noaman, A.Y.M.; Mousa Fardoun, H.; Ventura, S. Early dropout prediction using data mining: A case study with high school students. *Expert Syst.* 2016, 33, 107–124. [CrossRef]
- Chen, L.; Wang, L.; Zhou, Y. Research on Data Mining Combination Model Analysis and Performance Prediction Based on Students' Behavior Characteristics. *Math. Probl. Eng.* 2022, 2022, 7403037. [CrossRef]
- Jovinius, J. An Investigation of the Effect of Geographical Location of Schools to the Students' Academic Performance: A Case of Public Secondary Schools in Muleba District. Ph.D. Thesis, The Open University of Tanzania, Dar es Salaam, Tanzania, 2015.
- 40. Irigoyen, M.M.; Kurth, R.J.; Schmidt, H.J. Learning primary care in medical school: Does specialty or geographic location of the teaching site make a difference? *Am. J. Med.* **1999**, *106*, 561–564. [CrossRef]
- 41. Thiele, T.; Singleton, A.; Pope, D.; Stanistreet, D. Predicting students' academic performance based on school and socio-demographic characteristics. *Stud. High. Educ.* **2016**, *41*, 1424–1446. [CrossRef]
- 42. Sullivan, K.; Perry, L.B.; McConney, A. How do school resources and academic performance differ across Australia's rural, regional and metropolitan communities? *Aust. Educ. Res.* **2013**, *40*, 353–372. [CrossRef]
- Cresswell, J.; Underwood, C. Location, Location, Location: Implications of Geographic Situation on Australian Student performance in PISA 2000. ACER Research Monographs. 2004, p. 2. Available online: https://core.ac.uk/download/pdf/3678 0197.pdf (accessed on 15 November 2020)
- 44. Panizzon, D. Impact of geographical location on student achievement: Unpacking the complexity of diversity. In *Diversity in Mathematics Education;* Springer: Cham, Switzerland 2015; pp. 41–61.
- 45. Chang, D.Y. Applications of the extent analysis method on fuzzy AHP. Eur. J. Oper. Res. 1996, 95, 649–655. [CrossRef]
- 46. Lee, S.; Seo, K.K. A hybrid multi-criteria decision-making model for a cloud service selection problem using BSC, fuzzy Delphi method and fuzzy AHP. *Wirel. Pers. Commun.* **2016**, *86*, 57–75. [CrossRef]
- Murray, T.J.; Pipino, L.L.; Van Gigch, J.P. A pilot study of fuzzy set modification of Delphi. Hum. Syst. Manag. 1985, 5, 76–80. [CrossRef]
- 48. Ishikawa, A.; Amagasa, M.; Shiga, T.; Tomizawa, G.; Tatsuta, R.; Mieno, H. The max-min Delphi method and fuzzy Delphi method via fuzzy integration. *Fuzzy Sets Syst.* **1993**, *55*, 241–253. [CrossRef]