Specifying Allow the Student to Enter the Classroom Using Data Ming and the K-NN Algorithm

Serri Ismael Hamad
College of Education for Pure Sciences, University of Thi-Qar, Iraq.
E-mail: serriismael@utq.edu.iq

Wafaa Razzaq Hashim
College of Nursing, University of Thi-Qar, Iraq.
E-mail: wafaa_razzaq@utq.edu.iq

Received September 15, 2021; Accepted December 14, 2021
ISSN: 1735-188X
DOI: 10.14704/WEB/V19I1/WEB19197

Abstract

A classroom is a learning space in which both children and adults learn. Classrooms can be found at a variety of educational institutions, from preschools to colleges, as well as other venues that provide instruction or training, such as corporations and humanitarian and religious organizations. The classroom provides a place where students may learn without being distracted by outside factors. The subject instructor or the class administrator is the one who decides who is allowed to enter the classroom. We describe an approach in this paper to assist the teacher in determining which students are permitted to enter the classroom. We used one of our k-Nearest Neighbor classification approach is one of the most widely used mining algorithms, to predict which students would be permitted into the classroom depending on student data.

Keywords

Classroom, Data Mining, K-Nearest Neighbor (K-NN).

Introduction

Classroom techniques do have a significant impact on student accomplishment, and this influence is at least as powerful as with student background when combined with other characteristics of teaching understudy. Because of the overwhelming influence of their teachers' classroom activities, this research proves that schools do matter (Wenglinsky, 2002). The students are followers and listeners in the conventional teacher classroom. The flexibility to move around, initiate interaction, ask any questions, and establish restrictions on activity durations is provided from before the teacher. In addition, the teacher is also the one who provides the data and defines the key concepts (Muir-Herzig, 2004).

The classroom has made the process so easy and convenient that adds students directly, start up the class in minutes and generate classwork that appears on students' schedules, easily interact with guardians and simply give them updates by sharing a code or link with the entire class, Using Google Meet, which is incorporated into Classroom, can make face-to-face telecoms with students. Students can also submit their work using Classroom, which allows teachers to keep track with sections, evaluate grades quickly, and mark projects throughout any free time they get without the need to carry around stacks of paper. Thousands of apps approved by teachers work with Classroom to inspire creativity and provide limitless learning opportunities for connecting with every student (Allen, Joseph, et al., 2013). In a “real” classroom, students listen to the teacher's lecture, raise their hands to respond to questions, or operates independently on a written assignment. Technology can aid in the construction of knowledge in the classroom. Several experts believe that computers have an impact on teaching-learning activities. They claim that using computers in the classroom will make schools increasingly student-centered and allow for more individualized learning than ever before. Students can interact, apply critical thinking, and identify alternative solutions to challenges in today's student-centered classrooms with the help of computers (Muir-Herzig, 2004).

But in the classroom system, no person is allowed to enter the class, but rather each teacher must specify the persons allowed to enter his class, so the problem that arises is how the teacher can determine the students allowed to enter his classroom.

In this paper, we propose a method that helps the teacher determine which students are allowed to enter the classroom by mining student data and categorize whether he is allowed to enter or not to the class using the k-NN classification algorithm.

Background

Data Mining in Brief

Data mining is the process of extracting knowledge from massive amounts of data. Clusters or Classification, Association Rule Mining, and Sequence Analysis are the three basic components of data mining. Classification/clustering is defined as the process of analyzing a group of data and generating a group of grouping rules that may be used to categorize data for future data. The process of extracting knowledge from a data group and transforming it into a comprehensible structure is known as data mining. It is a computational method that involves approaches from machine learning, artificial intelligence, statistics, and database systems to uncover patterns in massive data sets (Kesavaraj, Gopalan, 2013).

1. Classification of Data Mining System

Data mining systems can be categorized using different criteria, specified these criteria as a type of database mined, type of knowledge mining, and kind of technique used and according to a kind of application adapted. Also, numerous strategies can be used in student data mining projects; the challenge is determining which methodology is best relevant for the issue at hand (Adeniyi, David Adedayo, 2016). A thorough data mining system can use various strategies or an integrated methodology that combines the advantages of several different interventions, according to the authors. Different technique of data categorization includes decision tree classifiers, Bayesian classifiers, K-Nearest Neighbor classifiers, and rule-based classifiers (Kesavaraj, Gopalan, 2013). In our research, the K-NN classification algorithm was used.

2. K-NN for Classification

The K-NN algorithm is a way for classifying objects in pattern recognition based on the feature space's nearest training examples. K-NN is a sort of lazy learning or instance-based learning in which the function is only approximated local and each calculation is deferred till classification. When there is little or no previous knowledge of the data distribution, the K-NN is the most basic and simplest categorization algorithm. During learning, this rule simply keeps the complete training set and gives a class to each query based on the majority class from its k-NN in the training set. When K = 1, the Nearest Neighbor rule (NN) is the simplest type of K-NN (Imandoust, 2013). In this manner, each sample must be categorized similarly to the samples around it. As a result, if a sample's classification is unknown, it can be predicted by looking at the sample's classification of its closest neighbors. The distance between all samples in a training set and an unknown sample can be calculated provided an unknown sample and a training group. The distance with the few value corresponds to the sample from the training group closest to the unknown sample. As a result, the unknown sample can be categorized based on its nearest neighbor's categorization (Adeniyi, David Adedayo, 2016). Figure (1) explains the steps of the K-NN algorithm.

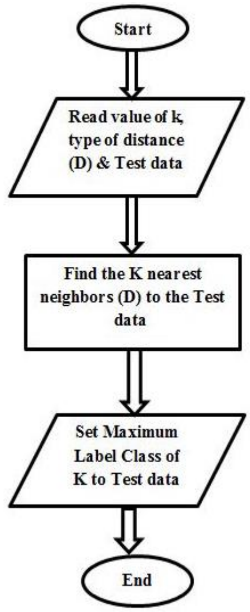

Figure 1 Steps of k-NN Classifier Algorithm

Methodology

This section delves into the implementation of the k-NN algorithm and how to use it to classify students and determine who is permitted to enter the classroom. The k-NN technique serves as a check belt around the resources we want to safeguard this is because only the most assured students will be permitted to cross over this barrier (see figure 2). Illustrates enter students to the classroom by using the K-NN algorithm; you can note that only highly reliable applicants can be entered with the small holes in the k-NN fence. These have a high level of reliability (the seven little black circles). Other requests (big circles) that do not meet the minimum level of trustworthiness are not considered.

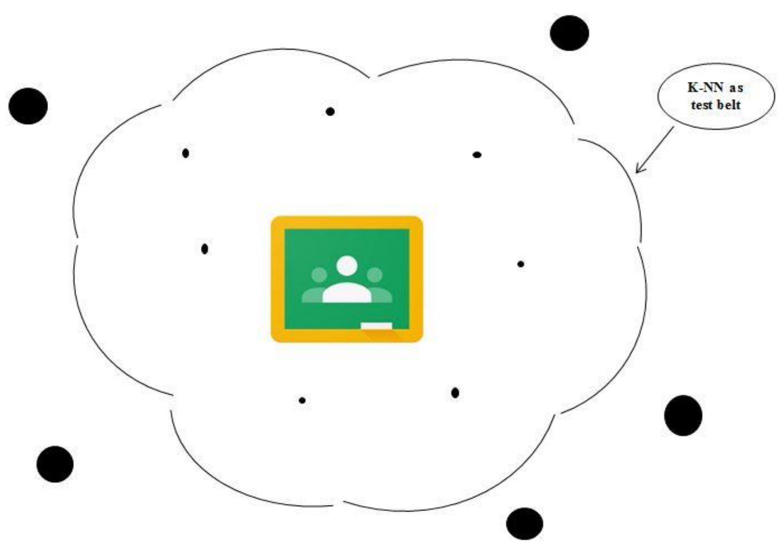

Figure 2 k-NN as test belt

Our Approach

Because the subject at hand is a classification issue, the K-NN data mining way is perfect. For class prediction, the K-NN method is the simplest and more straightforward (Adeniyi, David Adedayo, 2016). The system's goal is illustrated in the following scenario: Assume we have a separate class for a certain stage, such as the third stage, and a specific subject, such as artificial intelligence, for which there is, of course, a private teacher or administrator about this classroom. This teacher is in charge of allowing students into the classroom. To be admitted to this classroom, each student must possess a set of characteristics, which are as follows: (id, password, code of classroom, and student number), which enable him to enter the classroom.

Working of k-NN Classification Algorithm

The cosine similarity or Euclidean distance inter the training and test tuples are commonly used by the K-NN classifier, so the Euclidean distance method will be used to implement the K-NN approach for the purposes of this research (Adeniyi, David Adedayo, 2016).

Assume that our data tuples are limited to features of student are (id, password, code of classroom, and student number). The weighting conditions for each student are determined by the collection of characteristics that each student possesses. A test student expressed as a vector inside the space model is classified by the system by comparing it to all of the instances in it (i.e. the training group), with the similarity scale being the Euclidean distance similarity scale.

To determine a test student's classification, most of the same neighbors (i.e., training set) are observed together. Those with the same class Ci in their resemblance scores are added together to indicate the aggregate score of the Ci class among the test student's closest neighbors. The student is placed in the category that has the greatest overall resemblance.

The following formula can be used to calculate the Euclidean distances inter a training set and a test set:

Let F_i be an input tuple with q features (f_{i1}, f_{i2}, . . ., f_{iq})

Let a be the total number of input tuples (i =1, 2, . . ., a)

Let q be the total number of features (j =1, 2,. . ., q)

The Euclidean distance between tuple F_i and F_t (t =1, 2, . . ., a) can be defined as

$$(f_i \ , f_t) = \sqrt{(f_{i1} - f_{t1})^2 + (f_{i2} - f_{t2})^2 + \cdots + (f_{iq} - f_{tq})^2} \qquad (3.1)$$

Generally, the Euclidean distance between two groups for example

$F_1 = (f_{11}, f_{12}, \ldots, f_{1a})$ and $F_2 = (f_{21}, f_{22}, \ldots, f_{2a})$ will be,

$$dist(f_1 \ , f_2 \ ) = \sqrt{\sum_{i=1}^{a} (f_{1i}} - f_{2i})^2 \qquad (3.2)$$

All nearby points that are closest to the test set are contained in k-NN classification, and recommendations are produced based on the closest distance to the test set; this can be stated in the following way:

Let C be the expected class

$$Ci = \{f \in \mathrm{C}q; d(f, fi) \leq d\ (f, fw), i\#w\} \qquad (3.3)$$

When the k-NN classifier is used, it predicts the class that is labeled with the class C_i [5]. Actually, the value of k is frequently an odd number, such as 1, 3, 5, and so on. We simply applied the k-NN method and found that the experiment worked best with k= 1.

Application of K-NN Classification Technique to Predict Student's Class Labels in the Classroom

Example 1. Let us consider the data of students as a vector with four (4) attributes: (id, password, code of classroom, and student number), with students represented by $\mathrm{X1, X2, X3, X4, \ldots, X11}$ as displayed in table 1 as class labels. Assuming that student $\mathrm{X3'}$s class is unknown.

To find out the class of studentsX3, we have to calculate the Euclidean distance between the vector X3 and each all other vectors, by applying Eq. (3.2).

The Euclidean distance inter two groups for example training tuple X1 and test tuple X3 ie.

X1 = ($x11$, $x12$, $x13$, $x14$) and X3 = ($x31$, $x32$, $x33$, $x34$) each with the characteristics listed below as in table 1.

X1= (id, password, code of classroom, and student number), and X3 = (id, password, code of classroom, and student number), will be:

$$dist(x1\ ,x3\ ) = \sqrt{\sum_{i=4}^{n}(x1i - x3i\ )^2}$$

Table 1 the student data class labels training tuple

Student	Student number	Password	id	Code of classroom	Class
X1	1	64	134	610	accept
X2	2	91	107	184	accept
X4	4	81	198	443	accept
X5	5	34	158	668	accept
X6	6	14	87	468	accept
X7	7	25	59	638	refusal
X8	8	164	73	748	refusal
X9	9	171	137	546	refusal
X10	10	172	62	837	refusal
X11	11	20	111	832	refusal
X3	3	175	47	495	?

$$dist(x1\ ,x3\ ) = \sqrt{(x1{,}1 - x3{,}1)}^2 + \sqrt{(x1{,}2 - x3{,}2)}^2 + \sqrt{(x1{,}3 - x3{,}3)}^2 + \sqrt{(x1{,}4 - x3{,}4)}^2$$

This gives:

$$dist(x1\ ,x3\ ) = 115130.08931639027$$

Repeating the selfsame process in our instance for all other tuple X2, X4, . . ., X11, the result of these computations produced a stream of data as explain in table 2, which explained the students sorted by their Euclidean distance to the student X3 to be classified.

Table 2 Data showing students sorted by distance to student $x3$

Student	Class	Distance to student x_3
X_6	accept	26500.542371808166
X_4	accept	51756.32607710868
X_1	accept	115130.08931639027
X_7	refusal	143415.08095385227
X_5	accept	173267.09334146517
X_{11}	refusal	193789.43706507844
X_8	refusal	252827.3366983879
X_2	accept	311133.01909312035
X_9	refusal	337204.2879754052
X_{10}	refusal	342218.24407678796

The classifier K-NN predicts the class naming with class Ci for which $Ci = \{x \in Cp;\ d(x, xi) \leq d(x, xw), i \# w\}$. For the unknown student class i.e., the 1-NN classification simply selects the student who is closest to a student X3in terms of the least distance (ie. The first student from the top of the list), in table 2 for students $x3$ use their class labels for predict the class of X3, as a result, recommend similar headline of "accept" for student X3 as in student X6 class from table 2 as shown in figure 3. The k-NN algorithm is implemented and programmed using the Java Eclipse program.

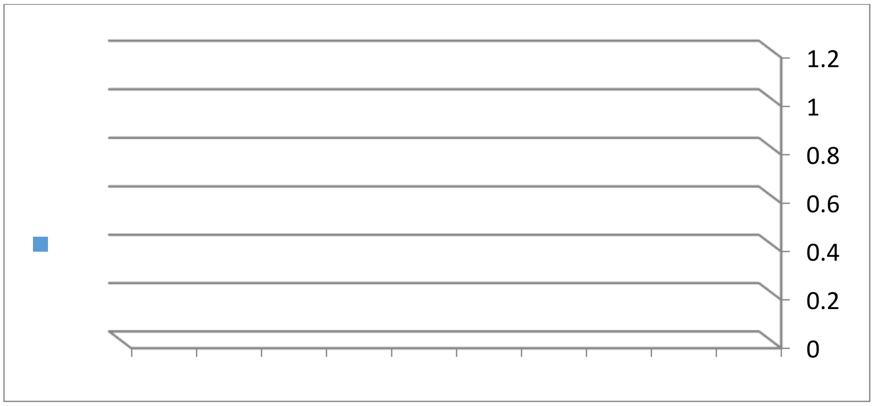

Figure 3 Distance to student x3

Conclusion and Future Work

We proposed a method in this research that assists the teacher in determining which students are permitted to attend the classroom by extracting student data and classifying whether or not he is permitted to enter. Where we used the simplest and most straightforward is the K-NN classification method to forecast the class in the classification and determine the students authorized to attend the class based on their characteristics, where each student must have a set of characteristics that allow him to do so. Our findings show that when K-NN classification is combined with the Euclidean distance method, it can provide helpful, very good, and accurate classification.

Propose to be executing the search by numerous other data mining techniques, comparison the result with this method, to determine model the more effective in handling an issue of this kind in the near future.

References

Wenglinsky, H. (2002). The link between teacher classroom practices and student academic performance. *Education policy analysis archives, 10*(12).

Muir-Herzig, R.G. (2004). Technology and its impact in the classroom. *Computers & Education, 42*(2), 111-131.

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of effective teacher–student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system—secondary. *School psychology review, 42*(1), 76-98.

Kesavaraj, G., & Sukumaran, S. (2013). A study on classification techniques in data mining. *In fourth international conference on computing, communications and networking technologies (ICCCNT)*, 1-7.

Adeniyi, D.A., Wei, Z., & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Applied Computing and Informatics, 12*(1), 90-108.

Imandoust, S.B., & Bolandraftar, M. (2013). Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. *International Journal of Engineering Research and Applications, 3*(5), 605-610.