Analysis Model for the Identification of Factors that Influence Students Academic Performance

Camilo Sandoval, David Tabla and Hector Florez*

Universidad Distrital Francisco Jose de Caldas, Bogota, Colombia

Abstract

Some studies identify different factors that influence academic performance. These factors are categorized into various dimensions such as academic, personal, economic, and social, among others. Then, it is important to choose the proper factor, when determining how academic performance might be affected. In this work, we propose an analysis model capable of relating factors with academic performance accompanied by an information-gathering tool that allows the parameterization of these factors and the visualization of a defined series of reports. Based on this model, it is possible to make different analyses and show if there is a significant influence of the factors on academic performance. For this, a series of factors focusing on a university undergraduate lecture are chosen in order to observe the operation of the model and therefore, offer input to make decisions to improve academic performance.

Keywords

Academic Performance, Analysis Model, Student Characterization, Data Analysis

1. Introduction

Studies such as those by Afzal et al., [1], Cruz et al., [2] and Tacilla et al., [3] identify different factors that influence academic performance. These studies made in different educational institutions conclude that there is a large series of factors that affect the academic performance of students. For example, Cruz et al., [2] carried out a study at the University of Puebla in Mexico, where they presented factors such as pregnancy, depression, family disintegration, stress, and mistrust, among others, which significantly affect the studies, as well as the development of their assigned tasks and work. In the same way, Afzal et al., [1] presents a work carried out at the University of Pakistan, where they consider that there is a significant influence of motivation on the academic performance of students.

As usual in the university context, there are undergraduate programs made up of different lectures that are taught throughout an academic period e.g., every semester. These lectures are focused on different subjects that are taught by different professors.

The proposed analysis model is developed on an online platform called DAS Web, where the information of the students, the lectures where students are enrolled, and the factors with which they are going to be related are compiled. These factors are attendance, participation,

D 0000-0002-5339-4459 (H. Florez)

CEUR Workshop Proceedings (CEUR-WS.org)

ICAIW 2023: Workshops at the 6th International Conference on Applied Informatics 2023, October 26–28, 2023, Guayaquil, Ecuador

[☆] casandovald@udistrital.edu.co (C. Sandoval); wdtablab@udistrital.edu.co (D. Tabla); haflorezf@udistrital.edu.co (H. Florez)

^{© 0 2023} Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

final grades, and results of surveys. In the DAS Web platform, a defined series of reports are exposed that relate to the factors analyzed by the model.

The importance of these reports is understood to the extent that the results identify population groups with certain factors that are related to their academic performance [4, 5]. So, for instance, students who are economically dependent on their parents can be categorized with a high or low grade, and also with other factors that are considered relevant such as means of transportation, time invested in getting to a lecture, attendance, participation, etc. Identifying this information makes it possible for an institution, a professor, or even a government entity to make decisions that improve academic performance.

The rest of the paper is structured as follows: Section 2 shows the definition of concepts related to the work. Section 3 presents a related word. Section 4 exposes the proposed analysis model supported in the DAS system. Section 5 presents the future work and finally, section 6 presents the conclusions.

2. Background

2.1. Biometrics

According to Jain et al., [6], biometrics is defined as the identification of individuals based on their biological or behavioral characteristics. These characteristics or behaviors can be of various types such as facial recognition, fingerprints, iris of the eye, dynamics of pressing the keys, and shape and style of writing, among many others. Thus, technology has been in charge of building means and tools that carry out the identification process through biometrics, for example, a fingerprint reader, which is a device capable of reading the fingerprint patterns on the human fingers in order to make recognition of an individual.

2.2. Academic performance

This concept has several nuances that have been nourished over time, but in the end, most of them are aligned to define academic performance as the result of learning and the product of an interaction at a didactic and pedagogical level between the two roles involved (professor and student). Tacilla et al., [3] provide different definitions such as the average of final grades and percentage of approval of a course, competencies in certain situations in the face of various academic stimuli, indicators and measures of the level of learning reached in the classroom by the student, among others.

Additionally, Tacilla et al., [3] expose academic, economic, and family dimensions and some determinants that somehow influence academic performance suggesting a relationship between the characteristics of a student and poor or good academic performance.

The quality of this academic performance can trigger different situations for a student. For institutions, low academic performance can cause obstacles to completing their studies if the student does not exceed the defined indicators. However, for high academic performance, institutions offer incentives and opportunities for a student, such as certifications, scholarships, and discounts, among others.

2.3. Student Characterization

Student characterization refers to determining the unique attributes of each student, attributes that can be contained in various dimensions. For instance, according to Alban et al., [7] there are a large number of attributes classified by dimensions such as personal dimension (age, gender, marital status, health problems, etc.), economic dimension (student employment situation, financial dependency, loans received, etc.), academic dimension (grades, load of academic credits, etc.), and social dimension (means of transportation, community support, political status, etc.). In this way it is observed that a student can be characterized by a large number of attributes and of different kinds; therefore, it will be a determining factor to select the characteristics that are desired to be obtained from the students to develop the desired analysis.

Another type of student characterization is behavior, for example, attendance and participation. In the first case, information is collected about the attendance of a student to a lecture, and whether or not the student actually attended the lecture. Information that can be captured by various means, such as attendance sheets or biometric readers integrated into a computer system. In the second case, the participation of the students in the classroom can be quite influential in its development, so capturing the information each time a student interacts or intervenes in the lecture provides new features for future analysis.

3. Related Work

A compilation of studies developed by different authors that relate aspects (student characterization, academic performance, analysis models) treated throughout the development of this research is presented below. For this, each work was carried out in different educational institutions and presents results that can support the academic processes of the students and the decision-making by administrative areas.

Cruz et al., [2] present research on the relationship between academic performance and some factors, mainly economic, of students from Puebla State University, Mexico. For this, a questionnaire of 7 closed questions and 1 open question was carried out, in order to investigate emotional, economic, social, family, and health aspects that affect academic performance. The sample consisted of 416 students from 11 groups, at different academic levels. From the results, it was possible to notice the relationship between the low academic performance and the economic insolvency of the students, added to low salaries that do not allow the acquisition of personal and school items, necessary for their studies.

Vásquez et al., [8] present various analysis models using data mining techniques to detect factors, related to the program and the individual context that may cause university students to drop out. They developed software where they analyze students who drop out and who do not drop out in the first 6 academic semesters. Based on 44 sociodemographic variables, academic performance, environment, and funding, obtained from each student, they applied different data mining techniques, concluding in determining the best models (clustering, no-balancing techniques, among others) and the most important variables. It was able to identify the most important predictors for each semester, and therefore, allows the identification of potential dropout students. Added to dropout prevention tools, a student can have a greater academic permanence.

Afzal et al., [1] conducted a study in different universities in Pakistan to identify the influence of motivation on the academic performance of students. To do this, they took a sample of 342 people to whom they conducted a survey with questions related to personal information, motivation (both extrinsic and intrinsic), and academic performance. From the results, it was concluded that intrinsically motivated students (self-exploration and altruism) perform better academically than extrinsically motivated students (rejection of alternative options, career and qualifications, social enjoyment, and social pressure).

Merchan et al., [9] present an investigation in which they formulated a predictive model of academic performance based on demographic data of first-year students at Universidad el Bosque. The demographic data that the model works with was provided by students as they are admitted to the academic program, and its academic performance data is provided by the weighted average of each student in the program. The information of 932 students is collected. From these, 465 students are used to feed the model and 467 to carry out the execution and validation of the model. Its methodology consists of two parts. The first part is executing a data mining process using the J48 and PART algorithms for the extraction of an initial set of rules used in the second part which is the formulation of the predictive model. Their results highlight that the socio-demographic factors that most influence the prediction of academic performance are marital status, day or night lectures, gender, number of siblings, and socioeconomic level. For the socioeconomic level, when it oscillates between a middle and high level, academic performance is favored. Finally, they conclude, among other aspects, that the prediction of academic performance reduces the possibility of dropping out. In addition, they propose that the institution department in charge begin to develop risk prevention strategies based on the socioeconomic classification of the new students in the program.

Alban et al., [7] present a work in which they propose factors that influence undergraduate university dropout according to the perspective of the students. The factors are supported by studies of 65 organizational and educational theories. To determine the causal relationship between the factors and university dropout, they use logistic regression and test it in a case study with 3773 students from a public university in Ecuador. The authors identify various factors based on their literature review and classify them into 5 dimensions: Personal, Institutional, Economic, Academic, and Social. Following this, they propose 11 new factors according to the perspective of the students such as planned and unplanned pregnancy, professors' commitment to the student, bullying, and sexism, among others. For the case study, they collected information through Google Forms surveys and performed an information analysis where they included the Box-Jenkins method to perform logistic regression. Their results indicate that factors such as the number of students' children and their perspectives on their integration into the labor market present higher dropout probabilities. When all the identified factors are present at the same time, the probability that a student decides to drop out of university is 95%. Finally, they conclude that the 11 identified factors have a strong causal relationship with dropouts, providing valuable information for decision-makers in university institutions.

Based on the presented works, it is observed how various models use factors of a social, academic, personal nature, etc., which can negatively or positively affect academic performance as well as being relevant when analyzing strategies for dropout prevention and permanence support.

4. Proposed Analysis Model

The proposed analysis model is depicted in the following subsections.

4.1. DAS System

The DAS system is the tool used to develop the proposed analysis model. Its composition is presented in Figure 1.

To introduce the proposed analysis model, a two-component approach is presented. These two components as *DAS Web* and *DAS BF*. Each of its logical elements is detailed relating the roles involved. Figure 2 presents the interaction between these two components.

4.1.1. DAS Web

It is a platform for the management of overall information (students, professors, lectures, visualization of attendance and participation, among other information). For greater understanding, it can be explained by dividing the platform into the following 4 categories:



Figure 1: DAS System Component Diagram.



Figure 2: Logic elements of DAS BF and DAS Web components.

- Overall academic management. Taking into account that the DAS system has an academic focus, it has concepts related to academic institutions, and more specifically, to universities. Therefore, it is important to know that there are professors who have assigned lecture groups in which they will offer a subject for each group in a certain period of time. The number of students allowed by the institution can be enrolled in a group of a lecture and in the same way, a student can be enrolled in different lecture groups. This information, in addition to other concepts, is illustrated in Figure 3. Each of these concepts can be managed by the administrator role of the system.
- Academic performance. Since academic performance can be defined in various ways, this project limits the concept to only the final grade assigned to every student by the professor.
- **Student characterization.** Knowing that each student has an endless number of characteristics of their own and in common, the present analysis model uses the results of a survey as one of the sources of information. To carry out this, the student role has a module called *Survey*, in which the student can answer the survey that must be previously defined by the administrator role. Initially, an administrator must create Perception Types (e.g., demographic, economic, psychological, etc.). Then, the administrator must create the questions that are categorized by perception type. Finally, the administrator must create the possible answers to those questions, which in the platform is found as *value*. Once the Survey is defined, a student can answer each question that will serve as input for subsequent reports.



Figure 3: Overall academic management.

• Attendance and participation. To talk about attendance and participation it is necessary to talk first about Timetables. The schedules determine the start time and the end time of classes in the lecture group. The start and end times are limited to integers. Having defined the schedules of a lecture group (part of the functions of an Administrator role), attendance and participation can now be allowed. Both can be registered through the other component of the system, which is DAS Biometric Fingerprint (DAS BF).

4.1.2. DAS BF

This component is the tool where students record attendance and participation in each of the lecture groups using their fingerprints. This component consists of a desktop application, tested on the Windows 10 Operating System, and a fingerprint reader device. This device must be installed prior to any registration.

4.1.3. Roles

The DAS system has two roles:

- **Student.** It is the actor who has characteristics displayed on DAS Web such as his final grade, enrolled lectures, data collected from the survey, attendance, and participation registered through DAS BF.
- Administrator. It is the actor that can parameterize most of the concepts in the DAS Web platform. His functions include:
 - Creation, edition, and deletion of students, professors, locations, lectures, and schedules, among others.

- Creation, editing, and deletion of perception types, questions, and possible values for each question.
- Visualization of the reports that can be generated through the collected data that will be the pillar of the proposed analysis model.

4.2. Analysis Model Flow

The analysis model flow is composed of 5 stages where stage 1 (Previous analysis) and stage 5 (Analysis of results) are not part of the scope of this work because these must be addressed under the analyst's criteria. The stages are presented in Figure 4 and explained as follows:

- **Previous analysis.** At this point, it is necessary to choose those characteristics of the student on which an analysis is going to be performed, i.e., determine the information that will be collected by each individual through a survey.
- **Parameterization.** At this point it is necessary for the administrator to parameterize the DAS system based on the following services:
 - Student registration on the DAS Web platform.
 - Registration of student fingerprints in the DAS BF software.



44

Figure 4: Analysis Model Flow.

- Registration of the parameters of the lecture group in the DAS Web platform.
- Registration of students to the lecture group.
- Survey registration. It includes:
 - * Types of perception: Configuration of the factors identified in the previous analysis which include demographic factors, economic factors, and academic factors.
 - * Questions: Configuration of the questions determined in the previous analysis.
 - * Question values: configuration of those possible answers to the previous questions.
- Gathering of information. It happens in different moments:
 - At the beginning of the academic period of the lecture, the student must enter the DAS Web platform and through a profile provided, fill out the survey parameterized in the previous point.
 - During the development of the lecture, the student must register their attendance through the biometric fingerprint reader, as well as their participation.
 - At the end of the academic period, the administrator must upload the final grade of the course for the enrolled students on the DAS Web platform.
- Execution of the model. Once all the necessary information is available, it is possible to launch the visualization of the model on the DAS Web platform through the Reports module, where various reports are available. It is only necessary to choose the type of report and press the button called Generate. The tables and graphs of the report are then displayed.
- Analysis of results. With the reports generated, the analyst can make his pertinent revisions, highlight whether the factors established in the previous analysis have a significant influence on the academic performance of each student, and in the same way relate academic performance with attendance and participation.

4.3. Analysis Model Reports

DAS Web offers a reports module that enables the deployment of different kinds of reports. For each report, there are two outputs: a table and a chart. All the tables have the possibility of being ordered in their columns and downloading their information in CSV files. Regarding the charts, the tool uses Plotly¹, which allows downloading the chart and slightly interacting with its results. It should be emphasized that the results described are supported by test data. The following are some reports offered by DAS Web:

- **Students attendance and participation.** This report relates grades with attendance and participation. It has the following outputs:
 - Table. It includes students' information, the total number of attendances, the total number of participations, and final grades.

¹https://plotly.com/



Figure 5: Report of students attendance and participation.

- Chart: It is a bar chart with a blue bar and an orange bar for each student, which corresponds to their attendance and participation in the group of the selected lecture. Additionally, next to the student's name, the value of the grade that the student obtained at the end of the course is displayed. Figure 5 presents the result ordered by final grade. It shows that the greatest number of attendances and participations belong to the students with the second highest grade in the group, suggesting a relationship between high academic performance and the fact to attend and participate in lectures. However, the rest of the students have similar data to each other (attendance and participation) without any influence on the final grade; therefore, those students with high academic performance can be considered as exceptions.
- Survey answers vs. total attendances and participations by lecture group This report relates one question with attendance and participation. For the question: *What means of transportation do you use to go to and from the university?*, it has the following results:
 - Table. It includes students' information, answers to the question, the total number of attendances, the total number of participation, and final grades.
 - Chart: It is a bar chart with a blue bar and an orange bar for each student, which corresponds to their attendance and participation grouped by answers to the question. Figure 6 presents the number of attendances and participation for each transportation means. For instance, students who move to the university by bicycle have 84 attendances and 68 participation. In the previous report, the hypothesis is that high academic performance is consequence of a greater attendance and participation. However, by relating the survey questions with attendance and participation, we can determine which factors can affect it.
- Late arrival attendance time in lecture group by schedule This report presents the number of late arrivals for each student during the period. It has the following results:
 - Table. It includes students' information, the total number of attendances, and final grades.



Figure 6: Report of students answers, attendance, and participation.



Figure 7: Report of late arrival.

- Chart: It is a series chart indicating the number of minutes that students arrive late every session of the lecture. Figure 7 presents the chart highlighting two students who report a late arrival of 24 and 44 minutes with grades of 47 and 25 respectively. However, using all data from the academic term, we can observe that the arrival time impacts academic performance.
- **Answers vs grades by lecture group** This report groups the grades by one question. It has the following results:
 - Table. It includes students' information, answers to the question, and final grades.
 - Chart: It is a box plot with different boxes based on the possible answers. Figure 8 presents the box plot relating socioeconomic level with final grades. It shows that the distribution of students is mainly of socioeconomic levels 2, 3, and 4; however, students with socioeconomic level 2 have the best average grades.
- Answers vs. late arrival attendance time in lecture group This report groups the grades by one question. It has the following results:
 - Table. It includes students' information, answers to the question, and final grades.
 - Chart: It is a box plot with different boxes based on the possible answers. Figure
 9 presents the box plot relating the time spent for transporting from home to the



What is your socioeconomic level?

Figure 8: Report of grades grouped by socioeconomic level.



How much time do you spend in total to transport from home to university?

Figure 9: Report of late arrival grouped by time spent in transportation.

university. In this example, this question is selected to analyze whether or not there exists a relation between the time spent in transportation and arrival time. To make a relation between this data and academic performance, it is necessary to analyze report number 3. For instance, report 3 suggests that arriving late produces a lesser academic performance, we can observe in this factor that a late arrival might be a consequence of several situations.

5. Future Work

The implementation of the analysis model in university students is a future work that is in progress. It is planned to be carried out in an undergraduate lecture at the Universidad Distrital Francisco José de Caldas and, for this, a prior analysis of the flow of the model is required i.e., to analyze the context of the institution and develop questions that characterize the students and contribute possible hypotheses regarding their academic performance.

At the end of the lecture, an analysis of the results obtained must be carried out to offer

stakeholders (professors, academic coordinators, etc.) a tool to identify those factors to work for different objectives such as student retention, lower academic desertion, and better academic performance.

6. Conclusions

Based on test data, the operation of the proposed analysis model is presented detailing each of its phases and components and supporting the importance of relating student characterization to academic performance, in such a way that it is understood the use of the application of the model.

Given the existing studies on different factors, it has been proven that they affect academic performance. However, not all factors influence the same way or in all contexts; therefore, the possibility of customizing how students will be characterized and the application of the analysis model provides stakeholders with more information to make decisions based on data.

This work includes 3 steps in the flow: parametrization, information gathering, and model execution excluding previous analyses and results analyses. On the one hand, based on the importance of the previous analyses, we suggest making a detailed analysis regarding which factors we would like to relate to academic performance taking into account attendance and participation as part of the students' characterization. On the other hand, the DAS Web platform offers the options to create, delete, and update the survey questions allowing different analyses when changing factors.

Finally, this work promotes the use of technologies for the strengthening of educational institutions and, likewise, for the development of society.

References

- H. Afzal, I. Ali, M. Aslam Khan, K. Hamid, A study of university students' motivation and its relationship with their academic performance, Available at SSRN 2899435 https: //ssrn.com/abstract=2899435 (2010).
- [2] E. R. Cruz Sosa, L. Gática Barrientos, P. E. García Castro, J. Hernández García, Academic performance, school desertion and emotional paradigm in university students, Contemporary Issues in Education Research (CIER) 3 (2010) 25–36.
- [3] I. Tacilla Cardenas, S. Vásquez Villanueva, E. E. Verde Avalos, E. Colque Díaz, Rendimiento académico: universo muy complejo para el quehacer pedagógico, Revista muro de la investigacion 5 (2020) 53–65.
- [4] K. Martinez-Zambrano, J. Paez, H. Florez, Pedagogical model to develop virtual learning objects for people with hearing impairment, CEUR Workshop Proceedings 2992 (2021).
- [5] O. Mendez, H. Florez, Applying the flipped classroom model using a vle for foreign languages learning, in: First International Conference on Applied Informatics, ICAI 2018, Springer, 2018, pp. 215–227.
- [6] A. Jain, R. Bolle, S. Pankanti, Introduction to biometrics, Springer, 1996.
- [7] M. Alban, D. Mauricio, Factors that influence undergraduate university desertion according

37 - 50

to students perspective, International Journal of Engineering and Technology 10 (2019) 1585–1602.

- [8] J. Vásquez, J. Miranda, Student desertion: What is and how can it be detected on time?, Data Science and Digital Business (2019) 263–283.
- [9] S. M. Merchan Rubiano, J. A. Duarte Garcia, Formulation of a predictive model for academic performance based on students' academic and demographic data, in: 2015 IEEE Frontiers in Education Conference (FIE), IEEE, 2015, pp. 1–7.