International Journal of Engineering Education Vol. 39, No. 5, pp. 1196–1206, 2023
Printed in Great Britain
0949-149X/91 \$3.00+0.00

Predicting Engineering Undergraduates Dropout: A Case Study in Chile*

CLAUDIA MARTÍNEZ-ARANEDA, MARIELLA GUTIÉRREZ VALENZUELA and MICHELLE BIZAMA VARAS
Facultad de Ingeniería, Universidad Católica de la Santísima Concepción. Alonso de Ribera 2850, Concepción, Chile.
E-mail: cmartinez@ucsc.cl, mariellag@ucsc.cl, mabizama@ucsc.cl

The main objective of this article is to present and validate a statistical model (N = 3,152) to predict the dropout of students from the School of Engineering of the Universidad Católica de la Santísima Concepción (UCSC) in Chile. Student droupout in engineering is a generalized and multifactorial phenomenon, even more so when the student can use his or her university access score for a period of two years. In the UCSC, a distinction is made between formal and nonformal droupout. The information collection methodology in this study included the survey administered by the Department of Evaluation, Measurement and Educational Registry of Chile (DEMRE) and input from the Directorate of Admission and Academic Registration of the UCSC. Within the analysis groups were students who formally resigned and were analyzed according to the reasons they gave for leaving; the other group was constituted by students who did not formalize their abandonment, deserters. Subsequently, a logistic regression analysis was applied to determine which variables would best explain the phenomenon of droupout. Among the main factors are gender (GENDER), program (AU), cumulative average score (PPA_SCORE), mathematics score of the university selection test (PSU_MATH_SCORE), mother education level (EDU_MOM), progression rate of student in engineering program (PROGRESSION_RATE) and socioeconomic quintile of student (QUINTILE). The performance of the prediction model shows an accuracy (88.53%) and precision (88.69%), which is a very encouraging result in relation to the performance of the studies reviewed in the literature.

Keywords: dropout; engineering; higher education

1. Introduction

Student droupout in engineering education is a generalized and multifactorial phenomenon, even more so since the Council of Rectors of Chile (CRUCh) resolved agreement 79/2010 in the plenary session N° 519 on October 28, 2010, held at Pontificia Universidad Católica de Chile, that the score of the University Selection Test (PSU) would remain valid for two years. This would allow students who take this test to apply to an engineering program and then, if they decide to do so, apply to another program or take the selection test again. The literature shows that the definitions of droupout are diverse: In [1] droupout was described as a *"voluntary or forced interruption of university studies that can occur temporarily or permanently, during the advancement of the curriculum without completing and achieving the degree of the training program"*. In [2] was defined as *"the premature abandonment of a study program before reaching the degree and time elapsed long enough to rule out the possibility of the student reincorporating"*. These definitions suggest both voluntary and nonvoluntary reasons for the phenomenon. In the Universidad Católica de la Santísima Concepción, a private institution with state support, a distinction is made between formal and nonformal droupout. The first includes all those cases in which the student voluntarily registers his or her abandonment, and in the second includes those cases in which this does not occur. The purpose of this work is to identify the predictive factors that affect student dropout, and for this purpose, both subsets are considered to align with the overall understanding of the concept.

The article is organized as follows: Section 2 characterizes the university, the faculty and the retention rates. Section 3 describes the theoretical approaches and models of droupout and a literature review on related works, and Section 4 details the work methodology to describe the main results obtained and validation of model in Section 5. Ending with the discussion of the results in Section 6 and the conclusions in Section 7.

2. Current Situation

2.1 Chilean Tertiary Education System

The Higher Education Information Service of the Ministry of Education of Chile (SIES) shows that by 2021, the Tertiary Education System of Chile was made up of 142 higher education institutions, of which 58 were universities, 34 were professional institutes and 50 were technical training centers. These institutions dictate undergraduate, graduate and postgraduate programs classified by area of

* Accepted 26 May 2023.

knowledge. In the same year, the system had 1,204,414 students enrolled in undergraduate programs, of which 57% were at universities, and of these, 22% were enrolled in technology studies (150,390), including Engineering, which is the object of this study.

First-year retention statistics in undergraduate programs show that in 2020, retention was 82.6% for engineering programs; 2.6% less in those related to information technology, and this area ranked among the 3 out of 10 areas with the most pronounced dropout rate.

Another important fact, which is related to this study, is that the retention rate of first-year students in professional studies shows that students from municipal secondary schools had worse retention rates than students from subsidized schools and other individuals, with these indicators being 80.3%, 83.9% and 87.2%, respectively, in 2020.

It is also important to note that in general, female students present better retention results for bachelor's degree programs in professional and technology studies.

However, if the indicator of expected curricular progress in professional courses of study by area of knowledge is analyzed, technology studies is again among the top 3 of 10 areas with the worst retention result, having a value of 80.6% for the year 2020, which was 3.4 percentage points below the average of 84.0%.

2.2 *UCSC School of Engineering*

The Universidad Católica de la Santísima Concepción (UCSC) was founded by the Archdiocese of Concepción in 1991 as a private, regional institution belonging to the Council of Rectors of Chile. The university serves approximately 12,000 students, of which 1,956 study in one of the five career paths offered by the Faculty of Engineering, which was established in 1998. Seventy percent of its students come from 40% of the most economically disadvantaged populations in Chile and are the first members of their families to access higher education. Due to inequities in school education, many of these students have deficiencies in their skills upon entry. The university understands and responds to this scenario by offering academic support, including student tutoring programs, workshops, and curricular activities through the Student Support Center (CEADE).

The School of Engineering offers psychoeducational support through the Psychoeducational Support Program (PAP) available to engineering students. This activity, together with the articulation of student services, is developed within the framework of a project with ministerial funds USC 1999, where one of the expected results is the decrease in dropout rates.

2.3 *Retention rates*

As observed in Table 1, first-year retention rates remain in the range of 74.75% to 81.46% for the period 2017–2022; in the case of second-year retention rates, there was an average decrease of 12.6% percent, and in the third year, especially in the fields of Electrical Civil Engineering, Geological Civil Engineering and Computer Civil Engineering, retention declines to 40%, which is partly explained by the fact that students can be expelled for academic reasons in the third year. The third-year retention rate in Industrial Civil Engineering was 74.56% and in Civil Engineering 66.02% in 2022 (source: Sentinel System, 2022). The retention behavior is shown in Table 1.

2.4 *Formal Dropout*

The UCSC distinguishes between RENOUNCED and DESERTED status, with the main difference being that students who renounce their course of study complete a form stating the causes for abandonment. As shown in Fig. 1, for the period 2015–2021, formal dropout numbered 557, and most of these students indicated they were changing institution (30.7%), other reasons (20.8%), vocational reasons (19.6%) and internal changes to a new program (7.5%). When categorizing of "other reasons", these were mainly attributed to vocational aspirations and family problems. Of the total

Table 1. Retention rates at the 1st, 2nd, and 3rd year in the Faculty of Engineering (2015–2022)

Year	2015		2016		2017		2018		2019		2020		2021		2022	
2014	341	76.29%	273	61.07%	243	54.36%	222	49.66%	201	44.97%	161	36.02%	106	23.71%	64	14.32%
2015	454	100%	340	74.89%	286	63%	269	59.25%	252	55.51%	239	52.64%	183	40.31%	105	23.13%
2016			490	100%	383	78.16%	294	60%	248	50.61%	232	47.35%	223	45.51%	170	34.69%
2017					424	100%	340	80.19%	265	62.5%	236	55.66%	229	54.01%	215	50.71%
2018							404	100%	302	74.75%	245	60.64%	223	55.2%	208	51.49%
2019									406	100%	294	72.41%	269	66.26%	243	59.85%
2020											348	100%	277	79.6%	220	63.22%
2021													329	100%	268	81.46%
2022															365	100%

Source: Sentinel System, 2022.

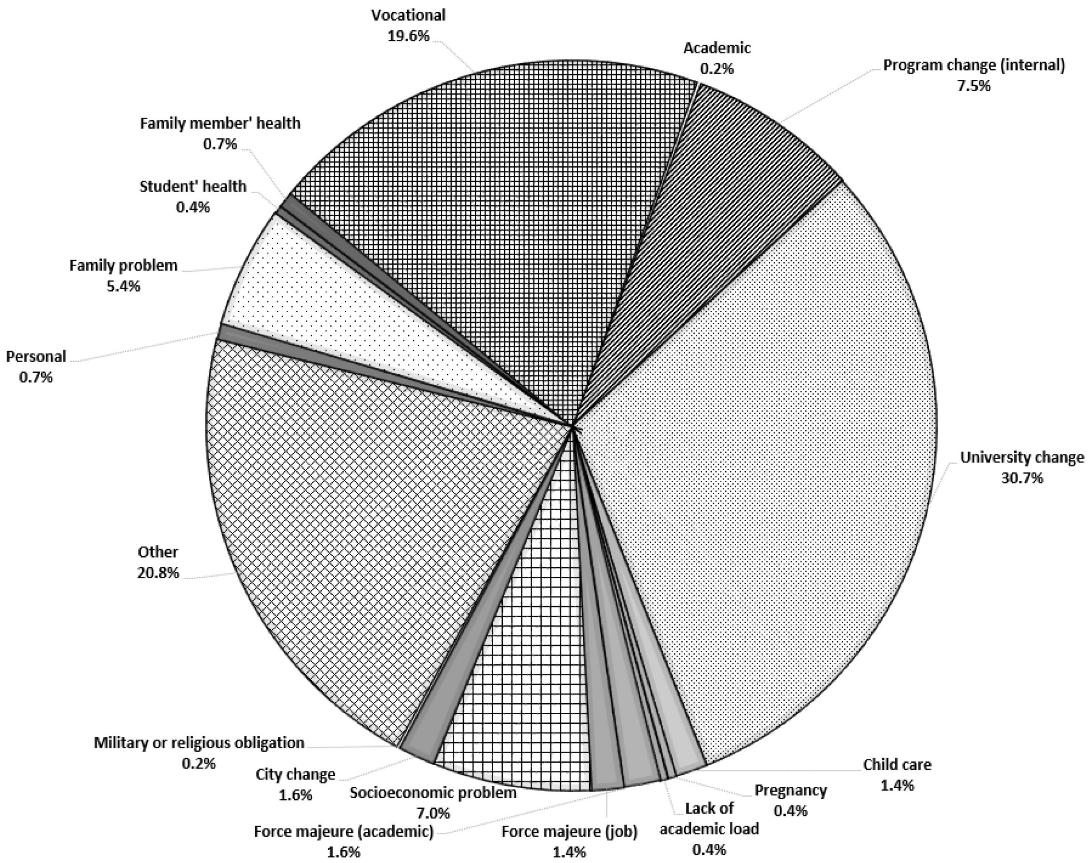

Fig. 1. Formal dropout causes (2015–2021).

number of resignations, 74.1% did so in the first year, 14.9% in the second year and 10.9% in the third year. For the purposes of this study, the total number of dropouts will be considered the union of the group of students who declared the cause and those who did not.

3. Background

The systematic review of the related works focuses on the theoretical models of dropout, in their dimensions and in the approaches through which the phenomenon is explained considering universities similar to the reality of UCSC.

3.1 Dropout Models

Within the theorical higher education dropout models are those of [2–4] in which the concept of academic and social integration is mentioned. According to the authors, academic integration occurs when there are situations of sharing academic values and social integration. These occur as a result of a student's relationships within the university community with his or his professors, peers, administrative staff members, and others. In general, terms, these models incorporate the individual, academic, institutional and socioeconomic dimensions, as shown in Fig. 2.

Depending on the theoretical model used to explain dropout, the psychological, sociological, and interactionist approaches appear (see Fig. 3).

The psychological approach is based on previous beliefs and behaviors; for example, previous academic performance influences the student's performance and self-concept [5, 9, 13]. The sociological approach explains dropout as a maladjustment in the integration of the environment and the family [3, 6]. The interactionist explains the permanence of higher education based on academic and social experiences [1, 4, 7, 8, 10–12, 14–17].

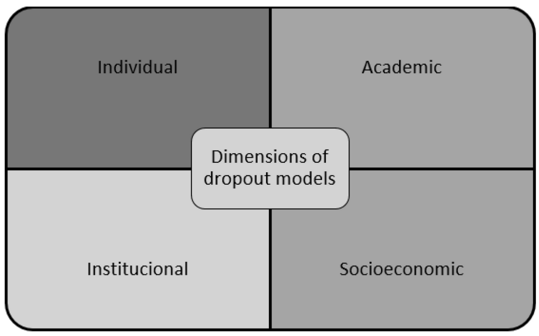

Fig. 2. Observable dimensions of dropout models.

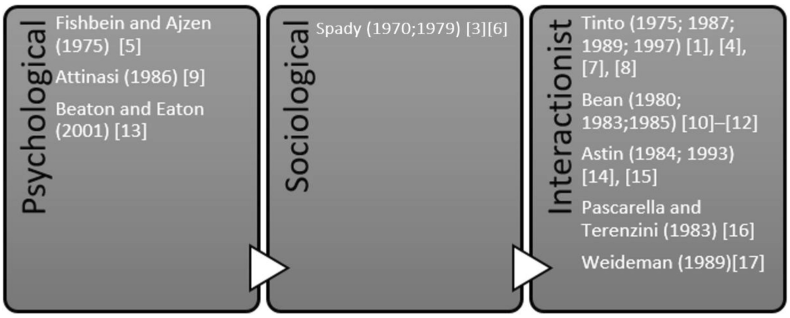

Fig. 3. Theoretical approaches from which dropout is explained.

3.2 Previous Dropout Studies

Table 2 summarizes the dropout studies carried out in the 2007–2021 period at universities similar to the UCSC in terms of student entry profile; in addition, studies that were not exclusively oriented to engineering, were included in their sample. The research topics (RT) considered in the literature review were the following:

RT1: Characteristics of the student sample.
RT2: Type of study(ies)/applied analysis technique(s).
RT3: Main results (risk/protective factors against dropout).

Table 2 shows that 3 of the 10 studies apply machine learning as technique to predict whether a student will dropout or not during their studies, the majority apply qualitative-quantitative and correlation analysis. Lack of motivation and poor performance of students in basic science courses are factors that are repeated in the results of the causes of the dropout problem.

4. Methodology

Fig. 4 shows the methodology for the generation of this study, the first stage includes a collection of data from formal and informal dropout students and the responses to a characterization survey administered to first-year students by the Department of Evaluation, Measurement and Educational Registry of Chile (DEMRE). With this dataset, we proceeded to the preprocessing stage, considering the cleanliness of the data based on completeness criteria. Next, in the stage of verification of variables entry conditions, their independence, normal distribution and noncollinearity were checked. Later, considering systematic reviews in which the logistic regression technique was mentioned [28], it was applied to these data until a minimum set of variables that would explain dropout was determined. The initial stages of the attrition model were developed with the PowerBI and Excel 365 tools and then continued with Python 3.10 and their statistical, visualization and machine learning packages such as Pandas, Numpy, Matplotlib, SciPy, Sikit-learn, and Seaborn.

5. Results

The results obtained for each stage of methodology are presented in the following sections.

5.1 Collection of Input Data

The input data corresponded to the records associated with all Engineering students in the period 2015–2022, including those who have graduated, those who remain regular students, and those who are in an intermediate state or temporary suspension, those who officially deserted and those lost to follow-up. There were 3,152 records, including personal, demographic, academic and family information, obtained from a characterization survey applied by DEMRE to higher education students each year. Each instance has 66 categorical and metric variables.

5.2 Data Pre-processing

The pre-processing of the input data began with the coding of categorical data such as GENDER, AU, EDU_MOM, EDU_TYPE, REGION, among others. For the completeness of the missing data, strategies were used that minimized the impact on the input data, for example, completing the missing data of REGION, according to the COMMUNE and completing the value of the PSU_AVG field with the value of the mean associated with the existing records. Completing the IVE (vulnerability indicator of the school of origin) field with information obtained from the Higher Education Information Service (SIES) and proceeding to the elimination of variables and instances that contained less than 50% completeness, less than 15% were affected, leaving 2,786 records available for analysis. It was found that the sample was balanced given that 58.2% corresponds to instances associated with student dropouts and 41.8% to non-dropouts.

Table 2. Relevant dropout studies

Authors	**Institution Country**	**TR1**	**TR2**	**TR3**
Canales and De los Ríos [18]	Chile	Forestry and Agriculture, Art and Architecture, Exact and Natural Sciences, Social Sciences, Law, Humanities, Education, Technology, Administration, and Commerce.	Qualitative-exploratory.	Vocational, motivational and socioeconomic.
Díaz [19]	UCSC, Chile	207 students from 3 Engineering programs	Quantitative based on Cox proportional risks.	Scholarship, Preference, College credit, Average university selection test, Come from a scientific-humanist school, Cumulative weighted average, Family income.
Saldaña and Barriga [20]	UCSC, Chile	329 students	Longitudinal explanatory study. Binary logistic regression.	Gross family income, Commune of residence, PSU Mathematics Score, PSU Language Score, Application preference, PPA_SCORE, Accumulated approved credits, Percentage of financing.
Argote, Jiménez, and Gómez [21]	Mariana University of San Juan de Pasto, Colombia	Student, cohorts 2005–2010.	Quasi-experimental study approach Quantitative.	Academic faculty, Financing, Type of institution, Gender.
Bernardo, Cerezo, Rodríguez-Muñiz, Núñez, Tuero, and Esteban [22]	Oviedo University, Spain	1,055 student cohorts 2010-2011.	Correlation study	Performance previous academic late enrollment, class attendance.
Carvajal and Cervantes [23]	Chile	10 evening students in Civil Engineering, Aquaculture Engineering and Commercial Engineering.	Qualitative study exploratory.	Conditions and personal and situational characteristics, capital and academic performance, unforeseen events and adverse circumstances, experience with institutional supply.
Alvarez, Callejas and Griol [24]	University of Computer Science, Cuba	456 students, year 2013–2014	J48 decision tree and a multilayer perceptron (MLP).	96.6% accuracy considering features of first year con MLP.
Lacave and Molina [25]	University of Castilla-La Mancha, Spain	363 students; 4 courses Computer Science Program	Bayesian networks.	Model not adjusted accurately.
Salas-Morera et al. [26]	Polytechnic School of Cordoba University, Spain	315 students Mechanical Engineering, 86 (43 of the first year and 43 of the second); Electrical Engineering, 44 (12–32); Industrial Electronics Engineering, 71 (37–34); and Software Engineering, 114 (60–54).	Quantitative and qualitative.	Students' lack of motivation, bad planning of the course by the students, high level of the course's starting point, syllabus too long, too many targeted activities, exams too difficult, inadequate class timetables, and inadequate examinations calendars.
Pinheiro et al. [27]	Polytechnic School at the Federal University of Bahia, Brazil	4,848 students, cohorts 2008–2016 Civil Engineering (1577), Mining Engineering (444), Electrical Engineering (786), Mechanical Engineering (796), Chemical Engineering (808) and Sanitary and Environmental Engineering (437 students).	Longitudinal study, survival analysis and correlation study.	Fail rates in Analytical Geometry, Calculus I, Calculus II and General Physics show highest correlation to student's dropout rates.

Fig. 4. Methodology for the generation of the study.

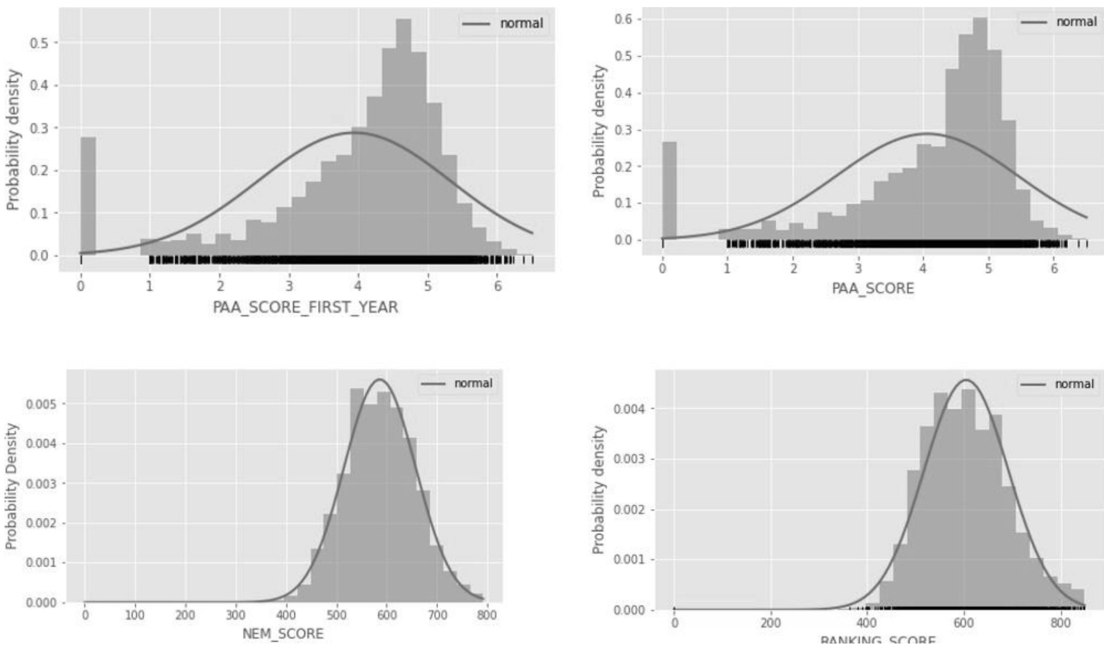

Fig. 5. Histograms with normal fit for some continuous variables.

5.3 Verification of Entry Conditions

Before applying the statistical model, it was necessary to complete verification of the variables' input conditions: (i) their independence, (ii) their normal distribution (continuous), and (iii) their noncollinearity. For the first condition, the chi-square test of independence al 95% was applied where in the case of categorical variables it was necessary to summarize the data in a contingency table, which allowed to eliminate variables containing the same information such as DESC_QUINTILE related to QUINTILE, ADMISSION_SUB_TYPE related to ADMISSION_TYPE, NAME_OF_PROGRAM related to AU, and PROGRAM_CRED determined by AU + PROGRAM; for the second condition, histograms were generated for the continuous variables (see Fig. 5), and the kurtosis and skewness statistics were calculated, which eliminated the variable IVE (see Table 3).

For the last condition, noncollinearity between quantitative variables was applied Pearson's correlation analysis [29, 30]. Fig. 6 shows that the greatest correlations between variables PROGRESSION_RATE and PASS_CRED (0.98), NEM_SCORE and RANKING_SCORE (0.97), PPA_RATE_FIRST_YEAR and PPA (0.94), CRED_REG_NEXT_YEAR and PROGRESSION_RATE (0.7), CRED_REG_NEXT_YEAR and PASS_CRED (0.7), so it was decided to eliminate PASS_CRED, NEM_SCORE, CRED_REG_NEXT_YEAR, and PPA_SCORE_FIRST_YEAR.

Finally, thirteen variables remained for analysis (see Table 4).

Table 3. Skewness and Kurtosis values for each continuous variable

Variable	Skewness	Kurtosis
PPA_SCORE	–1.593911506	2.071905847
PPA_FIRST_YEAR	–1.467066117	1.758830688
PSU_LYC	0.065552355	0.302744334
PSU_MATH	–0.561701517	1.235686056
PROGRESSION_RATE	0.83973077	–0.460603063
NEM_SCORE	–0.132147036	1.321946583
SCORE_RANKING	0.224538852	0.57703118
IVE	46.67188545	2361.200537

5.4 Binary Logistic Regression

Logistic regression is a special type of regression that is used to predict a categorical or binary variable [28]. Binary logistic regression analysis is applied when there is a variable that describes a response in dichotomous form (success = 1 or failure = 0), and it is desired to study the effect that other independent variables have on it, for example, in the case of estimating the probability that an engineering student with a certain entry profile will drop out or if one wishes to identify the predictability of a disease from a collection of symptoms. In general, terms, it is necessary to evaluate the influence that each independent variable has on the dependent variable. The general form of the binary logistic regression model is described in Equation (1).

$$P(Y = 1|X) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots\beta_n x_n}}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots\beta_n x_n}} \qquad (1)$$

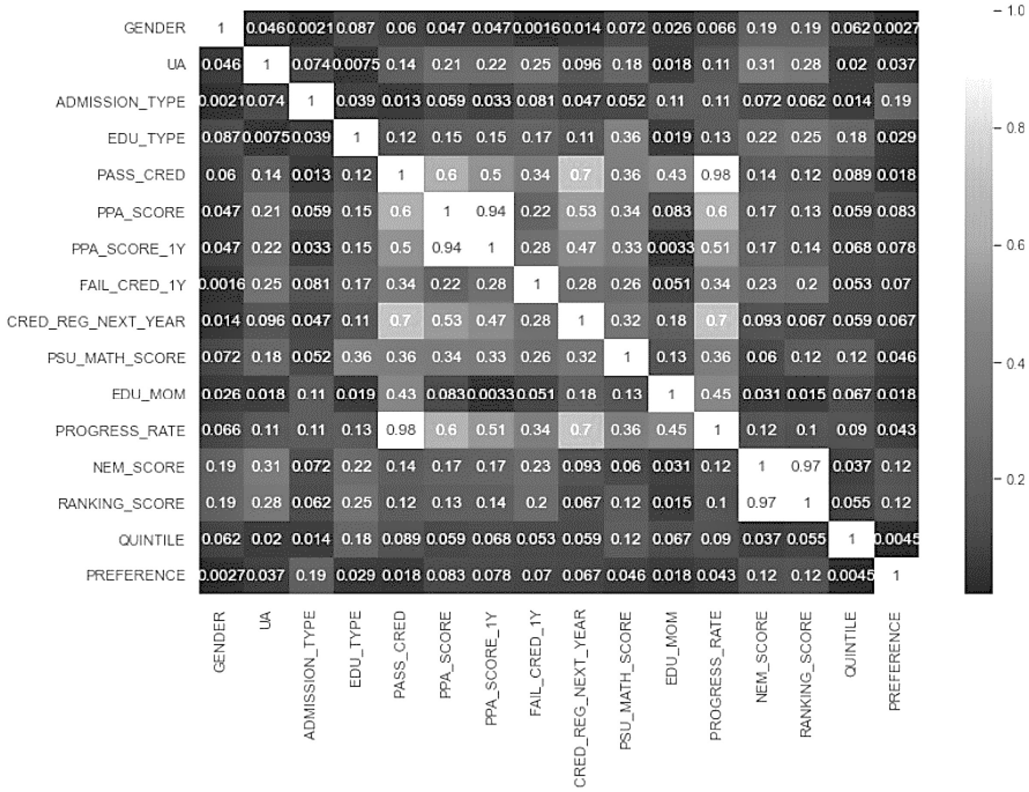

Fig. 6. Pearson correlation to between variables.

Table 4. Description of the input variables after preprocessing

Number	Variables	Type	Values
1	GENDER	Nominal	0: Male; 1: Female
2	EDU_TYPE	Nominal	0: Public; 1: Subsidized; 2: Private
3	UA	Ordinal	53:UA53; 54:UA54; 55:UA55; 117: UA117; 125:UA1225
4	ADMISSION_TYPE	Ordinal	0: Special; 1: Regular; 2: Direct
5	PPA_SCORE	Continuous	Range 1–7
6	FAIL_CRED_FIRST_YEAR	Discrete	0–30
7	PSU_MATH_SCORE	Continuous	302–738
8	EDU_MOM	Ordinal	–1: Unknown 0: Incomplete basic education 1: Basic education 2: Incomplete secondary education 3: Secondary education 4: Incomplete Technicians 5: Technicians and incomplete undergraduate program 6: Undergraduate program 7: Graduate program
9	PROGRESSION_RATE	Continuous	0–100
10	RANKING_SCORE	Continuous	Range 300–683
11	QUINTILE	Ordinal	1–5
12	PREFERENCE	Ordinal	1–10

where $x_1, x_2 \dots x_n$ are the independent variables, $P(Y = 1|X)$ is the probability of success of the dependent variable.

After fitting the model with the training data, the performance metrics observed in Table 5 were obtained. In turn, Table 5 shows those variables that have the lowest *p-value* (< 0.05), which would indicate that there is a significant effect on student dropout, which allowed to eliminate variables EDU_TYPE, ADMISSION_TYPE, FAIL_

Table 5. Predictor variables

Number	Variable	coef	p-value
0	Const	0.8265	0.4690
1	GENDER	0.6199	0.0000
2	UA	0.067	0.0060
3	PPA_SCORE	–0.1249	0.0000
4	PSU_MATH_SCORE	–0.7246	0.0040
5	EDU_MOM	–0.0462	0.0000
6	PROGRESSION_RATE	–0.1025	0.0000
7	QUINTILE	0.1597	0.0010

Table 6. Performance of the model

Accuracy	Precision	Recall	F1-Score
88.53%	88.69%	84.29%	86.44%

CRED_FIRST_YEAR, RANKING_SCORE and PREFERENCE.

Replacing values from Table 5 in Equation (2), the predictive equation would be:

$$logit(P(Y = 1|X)) = 0{,}8265 + 0{,}6199 * \text{GENDER} + 0{,}0067 * \text{UA} - 0{,}7246 * \text{PPA\_SCORE} + 0{,}0069 * \text{PSU}_{\text{MATH}} - 0{,}0462 * \text{MOTHER}_{\text{ED}} - 0{,}1025 * \text{PROGRESSION}_{\text{RATE}} + 0{,}1597 * \text{QUINTILE} \quad (2)$$

5.5 *Model Validation*

The model obtained was validated through a test set consisting of 20% (558) of the original instances that were not part of the training stage. The *accuracy* metric of the set of tests was 88.53%, which is very positive considering the proportion of the sample. The performance of the model can be seen in Table 6.

6. Discussion

In this section, the results are analyzed considering a set of test cases. Likewise, the generalization of the model to other Chilean and foreign institutions is discussed.

6.1 *Model Prediction*

As seen in Table 7, to determine the probability that a UCSC Engineering student with a certain entry profile will dropout, four arbitrary instances were defined, and the model applied.

From these test cases, with those variables with the greatest effect on the probability of droupout, it can be said:

(1) Test case 1: It is less likely that a student who has a higher PROGRESSION_RATE will drop out of the major, having a reasonable PPA_SCORE (5.5). In this case, decreasing the probability of droupout by 12.36% by increasing the rate of progression by 50%.

(2) Test case 2: If you change the GENDER variable from male (1) to female (0), decreasing droupout (under the same conditions). This could be based on the characteristics of women, and how they face adverse environmental factors described in [16] such as: changes at work, existence of people in charge, commitment to their family or the institution, among other. In this case, decreasing the probability of droupout by 15.32%.

(3) Test case 3: Related to the above, it can be seen that the EDU_MOM is an important factor when determining the probability of droupout, it is lower at a higher educational level. The explanation could prove from the fact that a family education within an environment that gives the student greater cultural capital can translate into better performance for him, a fact that is supported by the researchers [8]. Something similar is mentioned in the study by [19] where it is mentioned that when parents and support networks close to young people do not have experience of higher education, family and social orientation becomes weaker for the student. In this case, decreasing the probability of droupout by 8.96% by increasing the educational mother level by 3 units (see Table 4; row 9)

Table 7. Model predictions

Variable	Test case 1		Test case 2		Test Case 3		Test Case 4	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
GENDER	1	1	1	0	1	1	1	1
UA	125	125	125	125	125	125	125	125
PPA_SCORE	5.5	5.5	5.5	5.5	5.5	5.5	5.5	5.6
PSU_MATH_SCORE	567	567	567	567	567	567	567	567
EDU_MOM	4	4	4	4	2	5	4	4
PROGRESSION_RATE	10	15	10	10	35	35	35	35
QUINTILE	2	2	2	2	5	5	5	5
Logit	1.3205	0.8080	1.3205	0.7006	–0.6705	–0.8091	–0.7629	–0.83536
***P(dropout* = 1|*X*)**	**0.7892**	**0.6916**	**0.7892**	**0.6683**	**0.3383**	**0.3080**	**0.31801**	**0.30251**
	⇓	12.36%	⇓	15.32%	⇓	8.96%	⇓	4.88%

(4) Test case 4: Each tenth of increase in the PPA_SCORE decreases the probability of dropping out by 4.88%.

6.2 Generalization of Prediction Model

The system of access to universities in Chile is highly centralized, 79% of the universities are attached to the Single Admission System (SUA) (46 institutions for the 2024 admission process), which is administered by the Department of Evaluation, Measurement and Educational Record (DEMRE) dependent on the University of Chile [31]. In addition, there is the Higher Education Information Service (SIES), of the Chilean Ministry of Education, which as of 2017 requests High Education institutions information on program, enrollment, passing rates, retention, qualifications, among other data that allow the construction of general indicators for the entire system [32]. As mentioned in the methodology (section 4) and results (sections 5.1. and 5.2.), all the variables used in this study are taken from both the DEMRE and SIES databases, so the proposed model can be applied to the 46 Chilean universities that belong to the SUA.

On the other hand, there are studies that have compared the admission system to Chilean universities with that of other countries in the Americas and Europe. The study [33] analyzes the access systems of 10 countries, including Chile, through the degree of centralization and selectivity of the process, determining a great similarity between the SUA in Chile and SISU (Unified Selection System) in Brazil, which present a similar level of centralization and common characteristics such as a single platform, predetermined selection factors and distribution through the platform of applicants in available vacancies. In addition, Anísio Teixeira National Institute of Educational Studies and Research (INEP) of the Brazilian Ministry of Education manages a Census of Higher Education (CENSUP) that manages similar information to the SIES in Chile, where it is possible to know the progression of students, socioeconomic and family situation data [34]. This suggests that the prediction model could be applicable to higher education institutions in Brazil that report data to these national systems.

There are other countries, such as Argentina or Canada, which, according to [33], have highly decentralized admission systems; each university determines the selection criteria it will use and therefore does not apply a standardized admission test for all its institutions. In countries with these characteristics, this prediction model could not be applied.

In general, the proposed model has three groups of variables: (1) variables extracted from the SUA standardized test (PPA_SCORE, PSU_MATH_SCORE); (2) personal and socioeconomic variables (GENDER, EDU_MOM, QUINTIL) and (3) academic variables (UA, PROGRESSION_-RATE), therefore, it could be applied to any institution that applies a standardized test in the admission process and manages admission and academic data.

7. Conclusions

It was possible to characterize a group of students and the reasons for dropping out considering those who do so formally through the UCSC academic record system. For those who did not, the logistic regression model was adequate to predict student dropout considering mainly the variables GENDER, UA, PPA_SCORE, PSU_MATH_SCORE, EDU_MOM, PROGRESSION_RATE, and QUINTILE. These variables reflect in some way the individual, academic, institutional, and socioeconomic dimensions present in the theoretical models. The results were positively comparable with the performance of the models reviewed in the literature.

Within the limitations of the study, it is recognized, as expressed by the theoretical models, that socio-affective variables should be available to be added to the input variables that are currently related mainly to demographic, academic and family elements.

Likewise, the proposed prediction model is applicable to all higher education institutions in Chile and could be applicable to any countries which admission system to higher education institutions have a similar operation and data availability.

The study can be used to make certain types of decisions in relation to the identified predictor variables. As a background, the Psychoeducational Support Program of the Faculty of Engineering uses as predictor variables the grades of the first and second contest in critical courses to offer academic support from peer tutors. To improve the relevance of these actions, the variables identified in this study could be incorporated to make a decision to support those at risk of student dropout.

Acknowledgments – This study was carried out by the Directorate of the School of Engineering at UCSC within the framework of the FU 2019 USC1999 "Academic strengthening of the students of the Faculty of Engineering of UCSC, through psychoeducational support and the articulation of student services". We appreciate your willingness to support the processes of improving the quality of teaching in the Faculty of Engineering of the Universidad Católica de la Santísima Concepción.

References

1. V. Tinto, Defining Dropout: A Matter of Perspective, *Journal of Higher Education*, **71**(18), pp. 1–9, 1989.
2. E. Himmel, Models of analysis of Student Dropout in Higher Education, *Journal of Quality in Education*, **17**, pp. 75–90, 2002.
3. W. Spady, No Dropouts from higher education: toward an empirical model, *Interchange*, **19**(2), pp. 38–62, 1979.
4. V. Tinto, Dropout from Higher Education: A Theoretical Synthesis of Recent Research, *Review of Educational Research*, **45**, pp. 89–125, 1975.
5. M. Fishbein and I. Ajzen, Attitudes toward objects as predictors of simple and multiple behavioural criteria, *Psychological Review.*, **81**(1), pp. 59–74, 1975.
6. W. Spady, Dropouts from higher education: An interdisciplinary review and synthesis, *Interchange*, **19**(1), pp. 109–121, 1970.
7. V. Tinto, in University of Chicago Press, *Leaving College*: *Rethinking the causes and cures of student attrition*, 2nd edition, Chicago and London, 1987.
8. V. Tinto, Classrooms as communities: Exploring the educational character of student experience, *Journal of Higher Education*, **68**(6), pp. 599–623, 1997.
9. L. C. Attinasi, Getting in: Mexican American Students' perceptions of their college-going behavior with implications for their freshman year persistence in the University, *ASHE, 1986 Annual Meeting Paper.* San Antonio, TX, EE. UU, pp. 268–869, 1986.
10. J. Bean, Student attrition, Intentions and confidence, *Research in Higher Education*, **17**, pp. 291–320, 1980.
11. J. Bean, The application of model of turnover in work organizations to the student attrition process, *Review of Higher Education*, **6**(2), pp. 129–148, 1983.
12. J. Bean, Interaction effects based on class level in an explanatory model of college student dropout syndrome, *American Education Research Journal*, **22**(1), pp. 35–64, 1985.
13. J. Bean and S. Eaton, The psychology underlying successful retention practices, *Journal of College Student Retention Research,* **3**(1), pp. 73–89, 2001.
14. A. Astin, Student involvement: A developmental theory for higher education, *Journal of College Student Personnel*, **25**, pp. 297–308, 1984.
15. A. Astin, What Matters in College? Four Critical Years Revisited, *Journal of Higher Education*, **22**, pp. 74–75, 1993.
16. E. T. Pascarella and P. T. Terenzini, Predicting freshmen persistence and voluntary dropouts decisions from a theoretical model, *Journal of Higher Education*, **51**(1), pp. 60–75, 1980.
17. J. C. Weidman, Undergraduate socialization: A conceptual approach, in *Higher education: Handbook of theory and research*, J. C. Smar., pp. 189–322, 1989.
18. A. Canales and D. De los Ríos, Explanatory factors of university desertion, *Journal of Quality in Education*, **26**, pp. 173–201, 2007.
19. C. J. Díaz, Factors of student dropout in engineering: An application of duration models*, Technological Information*, **20**(5), pp. 129–145, 2009.
20. M. Saldaña and O. A. Barriga, Adaptation of Tinto's university dropout model to the Universidad Católica de la Santísima Concepción, *Journal of Social Science*, **16**(4), pp. 616–628, 2010.
21. I. Argote, R. Jiménez and J. Gómez, Detection of dropout patterns in the undergraduate programs of the Mariana University of San Juan de Pasto, applying the knowledge discovery process on a database (kdd) and its implementation in mathematical prediction models. *CLABES Conference*, Colombia, pp. 1–7, 2014.
22. A. Bernardo, R. Cerezo, L. Rodríguez-Muñiz, J. Núñez, E. Tuero and M. Esteban, Prediction of university dropout: explanatory variables and prevention measures, *Journal Sources*, **16**, pp. 63–84, 2015.
23. R. Carvajal and C. Cervantes, Approaches to university dropout in Chile, *Journal of Education and Research*, **44**, 2017.
24. N. Lázaro Alvarez, Z. Callejas and D. Griol, Predicting computer engineering students' dropout in Cuban higher education with pre-enrollment, *Journal of Technology and Science Education*, **10**(2), 241–258, 2020.
25. C. Lacave and A. I. Molina, Using bayesian networks for learning analytics in engineering education: A case study on computer science dropout at UCLM, *International Journal of Engineering Education*, **34**(3), pp. 879–894, 2018.
26. L. Salas-Morera, A. C. Molina, J. L. O. Olmedilla, L. García-Hernández and J. M. P. Romero, Factors affecting engineering student's dropout: A case study, *International Journal of Engineering Education*, **35**(1), pp. 156–167, 2019.
27. S. M. C. Pinheiro, K. Oliveira-Esquerre, M. A. Martins and R. Oliveira, Student Performance in First-Year Math and Physics Courses as Predictor of Student Dropout in Engineering Programs *International Journal of Engineering Education*, **37**(2), pp. 471–481, 2021.
28. M. Kumar, A. J. Singh and D. Handa, Literature survey on educational dropout prediction, *International Journal of Education and Management Engineering*, **7**(2), pp. 8–19, 2017.
29. C. B. Read and D. A. Belsley, Conditioning Diagnostics: Collinearity and Weak Data in Regression, 1991.
30. J. F. Hair, W. C. Black, B. J. Babin and R. E. Anderson, in Prentice Hall, *Multivariate Data Analysis*, 5th edn, Madrid 2013.
31. DEMRE. Admission process, https://demre.cl/proceso-admision/, accessed May 15, 2023.
32. SIES. Curricular Progress Reports in Higher Education in Chile, https://www.mifuturo.cl/informe-avance-curricular-en-educacion-superior/, accessed May 15, 2023.
33. Educar Chile, Comparative analysis of access systems to Higher Education, https://accioneducar.cl/analisis-comparado-de-los-sistemas-de-acceso-a-la-educacion-superior/, accessed May 12, 2023.
34. INEP, Census of Higher Education, https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior, accessed May 15 mayo, 2023.

Claudia Martínez-Araneda is an Associate Professor at UCSC, Chile. She was Head of School of Engineering from 2018 to 2023 and a faculty member in the Computer Science Department. Her research areas are Information retrieval, Sentiment analysis and Engineering education.

Mariella Gutiérrez Valenzuela is an Associate Professor at UCSC where she has been a member since 2000. Currently she is the College's Dean from 2016. Her research interests lie in the area of databases and geographic information systems. She has collaborated actively with researchers in several disciplines, applying the geographic information management and the development of applications for decision support to problems at the logistic sector, transportation, climate change effect in urban development, urban regulation, health and others.

Michelle Bizama is an Industrial Engineer at Universidad Católica de la Santísima Concepción, Chile and obtained her Master in Industrial Engineering at the Universidad Católica de la Santísima Concepción. Currently she is an assistant at the School of Engineering at UCSC, Chile. Her interest areas are statistical analysis and Engineering Education.