JOURNAL OF E-LEARNING AND KNOWLEDGE SOCIETY
Vol. 16, No. 01 (2020), pp. 44-54

Deep learning approach for predicting university dropout: a case study at Roma Tre University

Francesco Agrusti, Mauro Mezzini, Gianmarco Bonavolontà

[a] *Roma Tre University, Dept. of Education (Italy)*

(submitted: 29/11/2019; accepted: 13/02/2020; published: 30/04/2020)

Abstract

Based on current trends in graduation rates, 39% of today young adults on average across OECD countries are expected to complete tertiary-type A (university level) education during their lifetime. In 2017, an average of 10.6% of young people (aged 1824) in the EU-28 were early leavers from education and training. Therefore the level of dropout in the scenery of European education is one of the major issue to be faced in a near future. The main aim of the research is to predict, as early as possible, which student will dropout in the Higher Education (HE) context. The accurate knowledge of this information would allow one to effectively carry out targeted actions in order to limit the incidence of the phenomenon. The recent breakthrough on Neural Networks with the use of Convolutional Neural Networks (CNN) architectures has become disruptive in AI. By stacking together tens or hundreds of convolutional neural layers, a "deep" network structure is obtained, which has been proved very effective in producing high accuracy models. In this research the administrative data of about 6000 students enrolled from 2009 in the Department of Education at Roma Tre University had been used to train a Convolutional Neural Network based. Then, the trained network provides a predictive model that predicts whether the student will dropout. Furthermore, we compared the results obtained using deep learning models to the ones using Bayesian networks. The accuracy of the obtained deep learning models ranged from 67.1% for the first-year students up to 94.3% for the third-year students.

KEYWORDS: University Dropout, Deep Learning, Convolutional Neural Network, Educational Data Mining, Bayesian Network

DOI
https://doi.org/10.20368/1971-8829/1135192

CITE AS
Agrusti F., Mezzini M., Bonavolontà G., (2020), Deep learning approach for predicting university dropout: a case study at Roma Tre University.Journal of E-Learning and Knowledge Society, 16(1), 44-54. https://doi.org/10.20368/1971-8829/1135192

1. Introduction

Academic failure at Higher Education (HE) level can be divided into four main categories (Tanucci, 2006): a general irregularity in the achievement of credits / exams completion, the extended duration of the student condition (so-called *out-of-school education*), the lack of linearity of the career (e.g. course transfers) and finally the actual leaving of the learning path that leads to the exit from the university system without obtaining the degree. Of course, several are the variables that influence students' decision to leave university (Krause, 2005), and according to which prevails, even the dropout definition can vary, according to literature.

The definition chosen in this paper is the one proposed by Larsen and other researchers in 2013 where dropout is defined as "the withdrawal from a degree course before it is completed" (p. 18). This definition also includes withdrawal from individual courses of study but not students leaving due to pregnancy, illness, etc., i.e. for all those causes that can be attributed to very specific reasons and temporary duration. The phenomenon of university early leaving has several negative effects other than its consequences at a personal level. On a general level, low completion rates of a university course could lead to a bottleneck of the skills in a cohort of the population, that can have consequences on the economic and social level, decreasing the competitiveness, innovation and productivity of a country.

For decades, one of the most used and discussed models have been Tinto's "student integration" model, which underlines the importance of the academic and social integration of students in predicting the phenomenon of early school leaving. This model envisages five different approaches to integration: psychological, sociological, economic, organizational, and the interactionalist approach (Tinto, 1975, 2010). One of the other main models is the one proposed by Bean (Bean, 1988), the

”student attrition” model, based on the attitude-behavior of the student, which measures individual and institutional factors and evaluates their interactions to predict university dropout (Bentler & Speckart, 1979). Another interesting model of student/institution integration is the Pascarella model (Pascarella & Terenzini, 1980), which emphasizes the crucial importance of student success of having informal contacts with teachers. In other words, in this model, background characteristics interact with institutional factors influencing student satisfaction with the university. Numerous studies have demonstrated the positive effects of student-university interaction on persistence (Cox & Orehovec, 2007; Pascarella & Terenzini, 2005; Braxton, Shaw Sullivan, & Johnson, 1997). Event history modelling is another model much discussed in literature: proposed by Des Jardins, Albourg and Mccallan (DesJardins, Ahlburg, & McCall, 1999), this model takes into account the role of the succession of different events in the different stages of the student’s educational career, changing the importance of factors from year to year, depending on the time period.

In all these models, the relationship between students and institutions is relevant to reduce dropout rates (Cabrera, Castaneda, Nora, & Hengstler, 1992) and several strategies have been identified to improve student retention (Larsen, 2013; Siri, 2015).

From numerous U.S. research (Camara & Echternacht, 2000; S. Hu & Kuh, 2002; Kuncel, Hezlett, & Ones, 2004; Bridgeman, McCamley-Jenkins, & Ervin, 2000; Kuncel & Hezlett, 2007; Kuncel, Crede, & Thomas, 2007), the baccalaureate grade proved to be the best predictor of the performance of the first academic year (predicting better than the standardized SAT scores) and more specifically of the average grade that the student obtains at the first year of college (Perfetto, 2002). However, the link between the maturity grade and persistence in the educational system remains a controversial topic: Rosenbaum (Rosenbaum, 2004, p.2) asserts that ”the predictor of the probability that a student will graduate easier to use is still his grade of maturity”; likewise Ishitani (2006, p. 18) states that “the ranking position in the class at maturity has significant effects on the behavior of university attrition”. At the same time, however, other literature researches consider the maturity grade and the scores for standardized tests (e.g. SAT) insufficient to predict persistence at university (Ting & Robinson, 1998; Lohfink & Paulsen, 2005).

In Italy, due to the very high dropout rates in higher education (ANVUR, 2018), several specific studies were conducted (Burgalassi, Biasi, Capobianco, & Moretti, 2016; Moretti, Burgalassi, & Giuliani, 2017; Carbone & Piras, 1998) which confirmed the value of the baccalaureate vote (and of the entry skills of students more generally) together with the socio-demographic traits of the students (mostly the socio-economic context) as valid predictors of university dropout compared to the outcome of the first year of study.

Many of the models and studies carried out, both national and international, presented different analyses from the psychological point of view, building psychological-motivational models focused on expectation, reasons for involvement, personal value and motivation in general (Bandura, 1997; Bandura, Barbaranelli, Caprara, & Pastorelli, 2001; Marshall & Brown, 2004; Weiner, 1985; Gifford, Briceno-Perriott, & Mianzo, 2006; Covington, 2000; Pintrich, 2000). These models and surveys all involve the collection of data by interviewing students directly, through the use of tools (usually questionnaires) specially administered. The study presented in this article, however, aims to use only the data available in any university statistical office, without, therefore, at least at this stage of research, interviewing students directly. In this regard, it was decided to proceed to the analysis of these data through the use of Artificial Intelligence (AI).

Also literature show the use of different types of data analysis methodologies: correlational analysis (Araque, Roldan, & Salguero, 2009; Gutierrez et al., 2015; Bernardo Gutirrez, Esteban Garca, Gonzlez Garca, Nez Prez, & Dobarro Gonzlez, 2017; Willcoxson, 2010), univariate or multivariate variance analysis (Cukusic, Garaca, & Jadric, 2014), logistic regression and structural equations (Duque, Duque, & Suriach, 2013; Ghignoni, 2017; Santelices, Cataln, Kruger, & Horn, 2016), and multi-level analysis (Georg, 2009). The crucial issue with these statistic methodologies is the statistical requirements (i.e. data normality) and the difficulty of interpretation.

Today, AI and Machine Learning (ML) in general is used to replace human activities that are repetitive, for example, in the field of autonomous driving or for the task of classifying images. In these areas, AI competes with the man with quite satisfactory results and, in the case of abandonment of the educational system, it is extremely unlikely that an experienced teacher will be able to ”predict” the educational success of the student based on data provided by the administrative offices.

ML and statistical techniques have in common the main focus of learning the underlying phenomena through the analysis of previously generated data. However, they use two completely different approaches: ML algorithms need some requirements to be fulfilled but usually they are free from most of the statistical assumptions (i.e. a linear regression assumes a linear relationship between an independent and a dependent variable, independence of observations and homoscedasticity).

There are different approaches in ML: KNN and other lazy methods (Altman, 1992), tree construction-based methods (i.e. C4.5) (Quinlan, 2014), classification and regression trees (Breiman, 2001), Neural of Bayesian networks (Mitchell, 1997). The present study aims to learn about underlying phenomena of dropout of the full cohort of students in the Roma Tre University in Italy (R3U) by using ML-based methods to predict the phenomenon before it happens so to identify attrition paths and to prevent students' dropout by taking appropriate measures. The main aim of this study is to define a Convolutional Neural Networks (CNN) model (a particular type of Neural Networks (NN)) that can be used in other universities, defining and predicting dropout students' characteristics. We propose the following research hypotheses (RH):

- RH1 By using CNN it will be possible to predict student dropout for an entire cohort of students using only personal and non-academic characteristics
- RH2 By using CNN it will be possible to predict student dropout for an entire cohort of students on different degree courses using also academic characteristics

The recent advances on NN, made by using CNN, have been disruptive in the field of the AI. By stacking tens or hundreds of convolutional neural layers together, one gets a deep network structure, which has proven very effective in producing high precision models. These advances have shown that AI may be able to compete (or even exceed) with human capabilities in the tasks of classification and recognition.

The paper is organized as follows. In Section 2, are reported some of the most important studies on the use of AI for the prediction of university dropout. In Section 3, we briefly summarize the CNNs and Bayesian networks and we describe our custom CNN model. Then the metrics for the evaluation of these models are presented. Furthermore, it is described in detail a case study at Roma Tre University. In Section 4, we give some discussions on the methodology and on the results of the case study. Finally, in Section 5, conclusions on the study are briefly drawn.

2. Related literature

In this section, we discuss previous works that investigated the university dropout using Educational Data Mining techniques (EDM) (Bala and Ojha, 2012, Koedinger et al., 2015).

From the analysis of the literature it emerged that the algorithm of Decision Tree (DT) is the most commonly used for developing predictive models whose aim is to identify university dropout (Alban et al., 2019). A research conducted at the University of Chittagong examined the possibility of predicting university dropout using models based on the CART (Classification And Regression Tree) and CHAID (Chi-squared Automatic Interaction Detector) with the cross-validation folder to decide which model is more efficient than other in terms of accuracy (Mustafa et al., 2012). An Indian research has evaluated the models developed by DT algorithm using accuracy, precision, recall and F1 measure (Sivakumar et al., 2016). Another research project has implemented DT using socio-economic, academic and institutional data (Pereira et al., 2013).

In addition to DT, other classification methods were used in order to implement models for predicting university dropout. A research conducted at the University of Genoa used NN to detect students at risk of dropout (Siri, 2015). Another example is the work done at the College of Technology in Mato Grosso. The research presents a model developed with Fuzzy-ARTMAP Neural Network using only the enrolment data collected for seven-year period and the results show a success rate of accuracy over 85% (Martinho et al., 2013). In another study at Budapest University of Technology and Economics, 6 types of algorithms were employed to identify students at risk of dropout using the data of 15,285 university students (Nagy and Molontay, 2018).

The studies mentioned above use different data, algorithms, performance, metrics and methodologies therefore it is impossible to say which one is better than other due to heterogeneity. On the other hand, all the studies confirm the effectiveness of data mining approach to analyze and predict university dropout as highlighted by the work of Alban and Mauricio (Alban et al., 2019). As far as we are concerned (Agrusti et al., 2019, Mezzini et al. 2019), the difference from our approach and the above-mentioned studies is that we have used CNNs to analyze educational data.

3. Methods

The main aim of the research is to predict, as early as possible, which student will dropout in the HE context. In this research we want to investigate the use of deep learning and artificial neural networks for predicting the dropout of a student from HE. We further compare the results, obtained using the deep learning approach, with a more classical method which use Bayesian Networks.

3.1 Classification

One of the most important problem of the field of AI is the *classification problem* (LeCun et al., 2015). In the classification problem we have objects which can be images, sounds or written sentences and we want to associate to each object a class taken from a finite set K of predefined classes. If we represent each object as an n-dimensional vector of real numbers $\boldsymbol{x} \in \mathbb{R}^n$, the solution of the classification problem consists in finding a function $f : \mathbb{R}^n \rightarrow K$ that associates to each object $\boldsymbol{x}$ its class. We refer to $f(\boldsymbol{x})$ as the *true class* of $\boldsymbol{x}$.

3.1.1 Neural Networks

A NN can be viewed as a function ϕ that takes as input an n-dimensional vector $\boldsymbol{x}$ and produces a value, called *prediction* on $\boldsymbol{x}$. The prediction is *correct* when $\phi(\boldsymbol{x}) = f(\boldsymbol{x})$ and *incorrect* otherwise. Contrary to the classical process of algorithm design, in which the designer, in order to build the algorithm for solving a problem, needs a complete and thorough understanding of the nature of the problem to solve, (like for example in (Mezzini, 2018, Mezzini and Moscarini, 2016, Mezzini, 2016, Mezzini and Moscarini, 2015, Mezzini, 2012, Malvestuto et al., 2011, Mezzini, 2011, Mezzini and Moscarini, 2010, Mezzini, 2010, Mezzini, 2007)) in order to implement a NN, for the solution of a problem, the programmer can even be completely unaware of the mechanism or the semantic of the classification. For having a NN to produce correct predictions, we need that the NN undergo to a *training process*. The training process consists of feeding the NN with a set of objects, called the *training set*, and denoted as $T = (\boldsymbol{x}_i, f(\boldsymbol{x}_i)), i = 1, \ldots, N$ where N is the number of elements of the training sets. The class $f(\boldsymbol{x}_i)$, of each object $\boldsymbol{x}_i$ in the training set, is already known. This part of the training process is called the *forward pass*.

For each object $\boldsymbol{x}_i$ in the training set, the value $f(\boldsymbol{x}_i)$ is compared to the prediction $\phi(\boldsymbol{x}_i)$ of the NN. If the value of the prediction $\phi(\boldsymbol{x}_i)$ is different from its class $f(\boldsymbol{x}_i)$, the NN will be modified according to some optimization rule (Qian, 1999), in order to correct the error. Among different types of NN the CNN have gained much popularity since recently when cutting edge breakthrough have been obtained in the image classification task (Krizhevsky et al., 2012).

We employed three different architectures of CNN in order to test their effectiveness for our predictive model. The first two architectures, called respectively ResNetV2 (RNV2) (He et al., 2016) and InceptionResNetV4 (IRNV4) (Szegedy et al., 2017) represent the state of the art of CNN and perform the best or among the best (at the date of 2017) against industrial benchmarks. The third architecture, called DFSV1, was built by us by making modifications to the ResNet (He et al., 2016) and VGG architectures (Bengio and LeCun, 2015).

3.1.2 Bayesian Networks

Bayesian Networks (BN) are one of the most effective tool for the classification task (Pearl, 1988). Let $\boldsymbol{U} = \{A_1, \ldots, A_n\}$ be a set of discrete random variables. We call the set of all the possible different values the variable A_i can take, the *domain* of A_i. A BN describes a joint probability distribution of the set of random variables over $\boldsymbol{U}$ both qualitatively and quantitatively by using a directed acyclic graph (DAG) and a set of parameters. Formally a BN $\mathcal{B} = (G, \Theta)$ where G is a DAG whose vertex set is $\boldsymbol{U}$ and Θ contains the parameters of the network in the form $\Theta = \{\theta_A | A \in \boldsymbol{U}\}$ where $\theta_A = P(A|\Pi_A)$ where Π_A is the set of parents of A in G and $P(A|\Pi_A)$ represent the probability distribution of A given its parents Π_A. Based on this, we can decompose the joint probability distribution as

$$P(\boldsymbol{U}) = \prod_{A \in \boldsymbol{U}} P(A|\Pi_A)$$

For conducting all our tests with the BN we used the R package and the BN learning algorithms contained in *bnlearn* library (Scutari, 2010).

3.2 Model evaluation

Performance's evaluation of classification model is mainly based on the count of cases correctly and incorrectly classified.

Here we consider the case in which there are only two classes, labeled respectively $+1$ and -1.We represent the number of correctly and incorrectly classified cases in the *confusion matrix* (CM). On the CM the rows represent the real classes and the columns represent the predicted classes.

On a cell $(i, j), i, j \in \{-1, +1\}$ of the CM we put the number of cases predicted as *j* by the model but having real label *i* (see Table 1).

		Predicted	Class
		−1	+1
Real Class	−1	*TN*	*FP*
	+1	*FN*	*TP*

Table 1 - Confusion matrix of a binary classification problem.

Therefore, in the case of a binary classification problem, the CM is composed of four different values: true negatives (TN), false positives (FP), false negatives (FN) and true positives (TP) as reported in Table 1.

We use the CM to build several useful metrics in order to evaluate the performance of a classification model. The first metric to consider is the *Accuracy* that measures the ability of the classification model to provide reliable predictions on new data:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}}$$

When the number of real positive cases is much greater than the number of real false cases (or viceversa) the *accuracy* could be a misleading measure of the effectiveness of the model. Therefore, other evaluation metrics such as *precision* and *recall* are used. Specifically, *precision* is the proportion between true positives and all values classified as positive:

$$precision = \frac{TP}{TP + FP}$$

The higher the value of *precision*, the lower the number of FPs. On the other hand, the *recall* measure is the proportion between the true positives and all the values that are actually positive:

$$recall = \frac{TP}{TP + FN}$$

The *recall* describes how efficient is the model in recognizing the observed property. Furthermore, there is another metric that can be used to evaluate a classification model and it is called F_1 *measure*. This metric represents harmonic average between *recall* and

$$F_1 measure = \frac{2|TP|}{2|TP| + |FP| + |FN|}$$

The F_1 *measure* is useful for evaluating the model when *recall* and *precision* are equally important (Tan et al., 2005).

3.3 A case study at Roma Tre University

We collected, from the administration office of Roma Tre University, a dataset of students enrolled in the Department of Education (DE). The years of enrollment ranges from 2009 up to 2014 comprising a total of 6078 students. We found that 649 of all students were still active at the time when we acquired the dataset (August 2018), while the remaining 5429 closed the course of their studies either because they graduated or because they dropped out or by other reasons, explained later. We refer to this set of students as the *no active* students. Note that in the following when we will refer to the *enrollment year* (or simply the year) of a student we mean the number of years passed since her/his first enrollment to university, that is, we refer to an integer value between 0 and 9 since no student is enrolled for more than 9 years.

List 1	List 2
Year of beginning of studies	Academic year
Year of birth	Course code
Gender	Course name
Country of birth	Course year
High school type	Family income class
High school exit score	Working status
High school maximum exit score	Exemption from taxes
Year ending high school	Type of exemption from taxes
Transferred from other university	Handicap
ECTS from other university	Part time status
Faculty	Part time ECTS
	Type of renew of enrollment

Table 2 - List of administrative attributes.

3.3.1 Database construction

In general, each of the no active student is classified in two different classes: *Graduated* and *Dropout*. We excluded later all students which do not classified in these two classes, like for example students who changed faculty within the R3U or went to another university. The number of such students is 118. The number of graduated students is 2833 while the number of who dropped out is 2478.

We obtained, from the R3U's administrative office, most of the (out of what were available) administrative fields of all students. In the Table 2 is reported the list of administrative fields that are used.

Note that, for a given student, the value of the attributes in List 2 of Table 2, may change during her/his academic career from year to year, while the value of attributes in List 1 does not change during all her/his academic career.

List 3
Exam name
Score of the exam
Maximum score of the exam
ECTS of the exam
Exam date (month/day)
Academic year
Type of validation

Table 3 - List of the attributes relative to student's career.

The attributes in List 3 of Table 3 are relative to the student's academic career. They represent the attributes relative to each test or exam given by the student. Note that the field "ECTS of the exam" refer to the European Credit Transfer and Accumulation System.

In order to construct the training set all the domains of the dataset are converted, using an arbitrary bijective function, to a non-negative integer domain. For example, the domain of the attribute GENDER, was converted to the domain $\{0,1\}$ where 0 correspond to "male" and 1 to "female".

We created a table STUDENT, whose schema *S* contains all the attributes in List 1 of Table 2. For each field of List 2, we added to *S*, four fields, denoted as f_y where $y = 0,\ldots,3$, that is, one field for each of the first 3 year of enrollment to the university. We limited our tests only to the students that are still active at the year 3 because after that year the number of those students dropping out to university is very small and not significant from statistical and/or practical purposes. If a student ends his/her career in the year z, $0 \leq z < y$, then f_y will take the value δ for every year $z < y \leq 3$. The value of δ, which was arbitrarily chosen to be equal to -1, can be considered as a NULL value and it does not appear in the original domain of any field on the scheme *S*. Furthermore, for any field in the List 3 of Table 3 and for each year of enrollment $y \in \{1, 2, 3\}$ an integer m_y is set to represent the maximum number of exams sustained by any student on the year of enrollment y. We found that $m_1 = 24$, $m_2 = 19$ and $m_2 = 23$. Thus, for any field in List 3, for each year y and for each $z, 0 \leq z \leq m_y$, we added a field denoted as $g_{y,z}$. If a student in the year $y > 0$ of her/his academic career completes successfully no more than j exams, then the value of the field $g_{y,z}$ is set to δ for each $j < z \leq m_y$. Overall the table STUDENT has 530 fields (although we collected data up to year 5 totaling 897 fields).

			Validation								Test							
Year	T.	Arch.	Dropout		Degree		Acc.	Prec.	Recall	F1	Dropout		Degree		Acc.	Prec.	Recall	F1
			True	False	True	False					True	False	True	False				
0	B	RNV2	166	105	111	50	64,12%	61,25%	76,85%	68,17%	144	138	121	38	60,09%	51,06%	79,12%	62,07%
0	B	INCRV4	183	120	96	33	64,58%	60,40%	84,72%	70,52%	151	155	104	31	57,82%	49,35%	82,97%	61,89%
0	B	DFSV1	160	96	132	43	67,75%	62,50%	78,82%	69,72%	159	116	113	54	61,54%	57,82%	74,65%	65,16%
1	A	RNV2	65	10	228	20	90,71%	86,67%	76,47%	81,25%	44	16	199	37	82,09%	73,33%	54,32%	62,41%
1	A	INCRV4	67	17	221	18	89,16%	79,76%	78,82%	79,29%	50	23	192	31	81,76%	68,49%	61,73%	64,94%
1	A	DFSV1	65	22	216	20	87,00%	74,71%	76,47%	75,58%	52	26	189	29	81,42%	66,67%	64,20%	65,41%
1	B	RNV2	47	43	186	40	73,73%	52,22%	54,02%	53,11%	33	33	205	52	73,68%	50,00%	38,82%	43,71%
1	B	INCRV4	60	74	155	27	68,04%	44,78%	68,97%	54,30%	52	76	162	33	66,25%	40,63%	61,18%	48,83%
1	B	DFSV1	62	94	143	24	63,47%	39,74%	72,09%	51,24%	51	87	141	24	63,37%	36,96%	68,00%	47,89%
1	C	RNV2	61	23	215	24	85,45%	72,62%	71,76%	72,19%	44	25	190	37	79,05%	63,77%	54,32%	58,67%
1	C	INCRV4	54	24	233	17	87,50%	69,23%	76,06%	72,48%	42	34	195	43	75,48%	55,26%	49,41%	52,17%
1	C	DFSV1	54	28	229	17	86,28%	65,85%	76,06%	70,59%	45	30	199	40	77,71%	60,00%	52,94%	56,25%
2	A	RNV2	35	6	228	15	92,61%	85,37%	70,00%	76,92%	15	6	221	21	89,73%	71,43%	41,67%	52,63%
2	A	INCRV4	38	13	221	12	91,20%	74,51%	76,00%	75,25%	17	8	219	19	89,73%	68,00%	47,22%	55,74%
2	A	DFSV1	33	6	228	17	91,90%	84,62%	66,00%	74,16%	14	4	223	22	90,11%	77,78%	38,89%	51,85%
2	B	RNV2	16	9	243	15	91,52%	64,00%	51,61%	57,14%	11	14	213	41	80,29%	44,00%	21,15%	28,57%
2	B	INCRV4	15	5	247	16	92,58%	75,00%	48,39%	58,82%	12	13	214	40	81,00%	48,00%	23,08%	31,17%
2	B	DFSV1	17	10	242	14	91,52%	62,96%	54,84%	58,62%	15	21	206	37	79,21%	41,67%	28,85%	34,09%
2	C	RNV2	29	7	211	16	91,25%	80,56%	64,44%	71,60%	17	15	237	14	89,75%	53,13%	54,84%	53,97%
2	C	INCRV4	32	14	201	13	89,62%	69,57%	71,11%	70,33%	22	23	235	18	86,24%	48,89%	55,00%	51,76%
2	C	DFSV1	30	10	205	15	90,38%	75,00%	66,67%	70,59%	25	18	240	15	88,93%	58,14%	62,50%	60,24%
3	A	RNV2	19	3	94	6	92,62%	86,36%	76,00%	80,85%	13	11	91	10	83,20%	54,17%	56,52%	55,32%
3	A	INCRV4	19	1	96	6	94,26%	95,00%	76,00%	84,44%	13	4	98	10	88,80%	76,47%	56,52%	65,00%
3	A	DFSV1	20	3	94	5	93,44%	86,96%	80,00%	83,33%	12	3	99	11	88,80%	80,00%	52,17%	63,16%
3	B	RNV2	14	6	91	11	86,07%	70,00%	56,00%	62,22%	7	8	94	16	80,80%	46,67%	30,43%	36,84%
3	B	INCRV4	14	4	93	11	87,70%	77,78%	56,00%	65,12%	1	4	98	22	79,20%	20,00%	4,35%	7,14%
3	B	DFSV1	16	8	89	9	86,07%	66,67%	64,00%	65,31%	5	10	92	18	77,60%	33,33%	21,74%	26,32%
3	C	RNV2	17	3	94	8	90,98%	85,00%	68,00%	75,56%	11	5	97	12	86,40%	68,75%	47,83%	56,41%
3	C	INCRV4	18	6	106	4	92,54%	75,00%	81,82%	78,26%	19	8	105	14	84,93%	70,37%	57,58%	63,33%
3	C	DFSV1	17	2	95	8	91,80%	89,47%	68,00%	77,27%	12	5	97	11	87,20%	70,59%	52,17%	60,00%

Table 4 - Here we report the confusion matrix for the epochs with the best F_1 measure on the validation set. The confusion matrix for the test set was computed using the very same model that achieved the best F_1 measure on the validation set. Column 'T' stands for table type (A, B or C).

We build a table called Y_LABEL containing two attributes: STUDENTID and DROPOUT, where the last represents the label of each student. It has a numerical domain with the following meanings:

$$\text{DROPOUT} = \begin{cases} 0, & \text{if the student graduated} \\ 1, & \text{if the student dropped out} \end{cases}$$

From the table STUDENT described above, we derived three type of tables denoted as STUDENT_A$_x$, STUDENT_B$_x$ and STUDENT_C$_x$ for $0 \leq x \leq 3$ where x is the number of years from the first enrollment.

In the schema of tables STUDENT_A$_x$ we added all the attributes in List 1 and all the attributes in List 2 (of the type f_y), and all the attributes of List 3 (of type $g_{y,z}$) for all $y = 0,\ldots,x$.

The tables denoted as STUDENT_B$_x$, $x = 0,\ldots,3$, contain only the attributes of List 1 and List 2. That is, we considered in these tables only administrative fields and we excluded the fields related to the academic careers of the students (the ones of type $g_{y,z}$).

The tables STUDENT_C$_x$, $x = 0,\ldots,3$ have been constructed in the following way. We computed, for each student, the following aggregate statistics: DIFFYEAR and for each year $x > 0$, NUMBEREXAMS$_x$, AVGSCORE$_x$ and SUMETCS$_x$. The first statistic contains the value

YEAR OF BIRTH − YEAR OF BEGINNING OF STUDIES − 19

that is, the difference in years between the age of the student (at the date of the enrollment) and 19. The other statistics contains, for each student and for each year $x = 1,2,3$ respectively, the number of exams successfully passed, the average score of the exams successfully passed and the sum of the ECTS gained. We thus obtained the schema of STUDENT_C$_x$ by adding to the schema of each table STUDENT_B$_x$, all the above four fields. The idea we want to test here is whether it is better and effective to use only some significant aggregate statistics or, instead, it is better and effective to use all the attributes relative to the academic career (like in tables STUDENT_A$_x$).

For the tests of both CNN and BN we choose a random permutation of all no active students. Next, we partitioned all students in twelve different mutually disjoint groups containing approximately 450 students each thus obtaining a partition $\mathcal{P} = \{P_0, P_1, \ldots, P_{11}\}$. For all $0 \leq i \leq 11$ the group P_i is used as a validation set V_i and the group $P_{i+1\ mod\ 12}$ as a test set T_i and the students in the remaining groups, as the training set A_i.

In the validation set for the year x we put only the students who, at that year of enrollment, were still active.

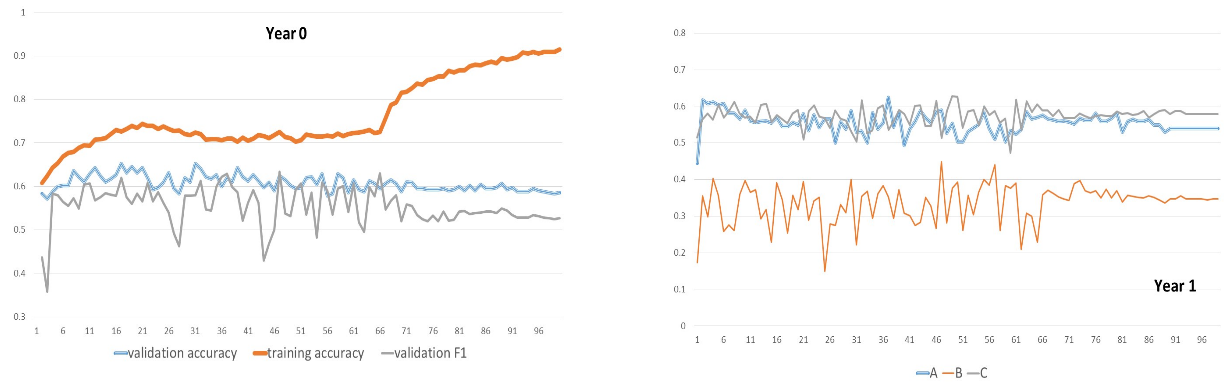

Figure 1 - On the left. Accuracy of the training of data taken from table STUDENT_B$_0$ for the validation set, the training set and the F_1 measure for the validation set. Students of the validation set are taken from the partition group 0. On the right. The F_1 measure for the validation set, year 1, partition group 0 and for the three different tables STUDENT_A$_X$, STUDENT_B$_X$ and STUDENT_C$_X$. For both figures the horizontal axis is the number of epochs.

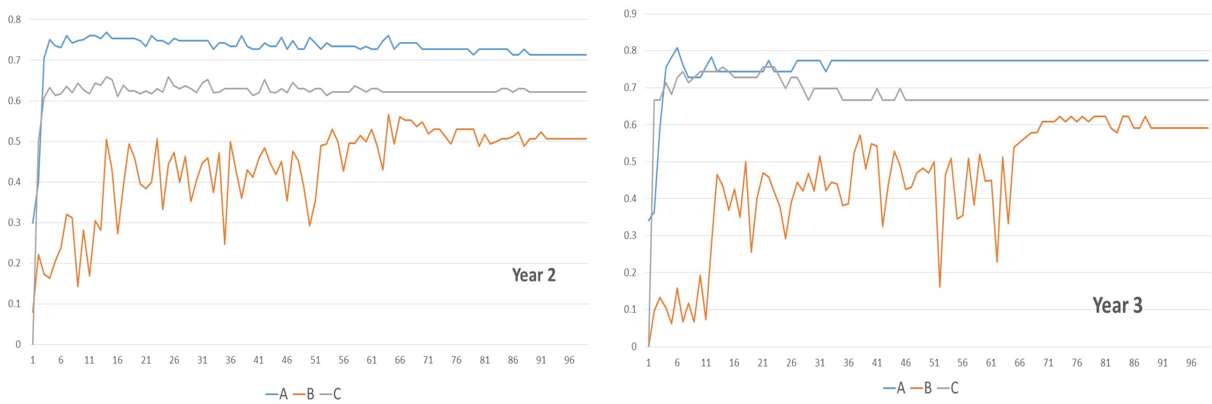

Figure 2 - The F_1 measure for the validation set, partition group 0 and for the three different tables STUDENT_A$_X$, STUDENT_B$_X$ and STUDENT_C$_X$. On the left. Year 2. On the right. Year 3. For both figures the horizontal axis is the number of epochs

3.3.2 CNN Training experiments and data

We trained three models based on the CNN architectures mentioned above by taking from each of the table above (A or B or C) the training, validation and test sets from the partition $\mathcal{P}$.

We got data from a total of 43200 epochs. For each epoch the confusion matrix of both the validation and the test sets were produced.

We found that the F_1 measure, was the better indicator for the selection of the best model.

In Figure 1 and Figure 2 we report the graph relative to the training of the model RNV2 for 100 epochs. In the graph on the left of Figure 1 we report the accuracy for the validation set, for the training set and the F_1 measure. Training and validation data where taken from the table STUDENT_B$_0$. In the graph on the right of Figure 1 and in the graphs of Figure 2 we report the value of the F_1 measure for the year 1 (Figure 1 on the right) and for the years 2 and 3 (Figure 2) for the three different tables STUDENT_A$_X$, STUDENT_B$_X$ and STUDENT_C$_X$. We observe that in all three cases the value of the F_1 measure relative at the table STUDENT_B$_X$ is always worse in every year. This clearly shows that using only administrative data gives very poor performance in predicting the dropout of a student. In Table 4 we report the data of the confusion matrix, for both validation and test sets, in which the validation set, among the twelve possible different sets of the partition $\mathcal{P}$, achieved the best score on the F_1measure.

3.3.3 Bayesian Networks training experiments and data

In order to compare the results and better understand the quality of the data produced by the CNN models, we executed extensive tests using BN on the same dataset. We used the *bnlearn* library available for the R package. In the bnlearn library there are several algorithms, for learning the BN from data, which are divided in three classes: (i) structural based learning, (ii) score-based learning and (iii) mixed structural and score-based learning. Furthermore, it has two classifiers, based on naive Bayes and tree Bayes (Friedman et al., 1997). In all the cases we used the default hyperparameters or default score functions for learning the model.

Year	Type	Algorithm	Validation								Test							
			Dropout		Degree		Accuracy	Precision	Recall	F1	Dropout		Degree		Accuracy	Precision	Recall	F1
			True	False	True	False					True	False	True	False				
0	B	14	118	45	171	98	66,90%	72,39%	54,63%	62,27%	102	67	192	80	66,67%	60,36%	56,04%	58,12%
1	A	9, 10	64	17	221	21	88,24%	79,01%	75,29%	77,11%	45	16	199	36	82,43%	73,77%	55,56%	63,38%
1	B	15	38	48	177	46	69,58%	44,19%	45,24%	44,71%	28	40	194	56	69,81%	41,18%	33,33%	36,84%
1	C	3, 5, 6, 9, 12	64	17	221	21	88,24%	79,01%	75,29%	77,11%	45	16	199	36	82,43%	73,77%	55,56%	63,38%
2	A	9, 10	35	13	220	15	90,11%	72,92%	70,00%	71,43%	18	15	212	18	87,45%	54,55%	50,00%	52,17%
2	B	2, 8, 9, 10, 13	37	19	215	13	88,73%	66,07%	74,00%	69,81%	20	22	205	16	85,55%	47,62%	55,56%	51,28%
2	C	5, 9, 10	32	9	225	18	90,49%	78,05%	64,00%	70,33%	17	2	225	18	92,37%	89,47%	48,57%	62,96%
3	A	9, 10	18	4	98	5	92,80%	81,82%	78,26%	80,00%	17	11	92	7	85,83%	60,71%	70,83%	65,38%
3	B	6,12	13	0	90	12	89,57%	100,00%	52,00%	68,42%	4	6	95	14	83,19%	40,00%	22,22%	28,57%
3	C	5, 9, 10	19	6	87	3	92,17%	76,00%	86,36%	80,85%	23	7	94	10	87,31%	76,67%	69,70%	73,02%

Table 5 - The confusion matrix for the BN models with the best F_1 measure among all different twelve groups.

We compute the predictions both on the test and the validation set and, following the same methodology used with the CNN, we considered only those models that give the best F_1 measure. The results are presented in Table 5. In the column "Algorithm" we reported the id of either the algorithm or the classifier that give the best result on the F_1 measure.

When there is more than one id this means that all the algorithms gave exactly the same results.

4. Discussions and Conclusions

We explored the effectiveness of predicting the dropout from university using three different sets of features. The first one, containing all the academic and administrative features (tables STUDENT_A_X). The second one, containing only administrative features (tables STUDENT_B_X) and the third (tables STUDENT_C_X) containing the administrative features and 3 aggregate statistics about the academic career of the students. The experiment showed that using only administrative features does not give good results and the models using only them are always outperformed by models using also the academic career features or aggregate statistics.

Furthermore, the models using, besides administrative features, also aggregate statistics perform slightly worse than the models using only and all the academic careers features. From all the above discussion we clearly conclude that the more accurate data we have the more precise and effective the model's predictions could be. The experiments done also demonstrated that using CNNs give us better results than using the BNs.

We implemented several state-of-the-art CNNs models, using real data of students of the DE in the R3U enrolled between 2009 and 2014. We also developed several BN models for predicting the university dropout. We compared the experiments made with CNNs with the one using the BNs. Much works could be also developed in the future. First, we can incorporate in the data the fields not included due to privacy censoring. We tested only three different architectures, but many other different CNN architectures exist in literature (Hu et al., 2018). Furthermore, many of the parameters of these architecture could be modified, and much could be explored in order to increase the effectiveness of the models.

Since it is not required that the prediction process is made in real time, we can train hundreds of models and make multiple prediction in order to reduce the random variation found in the early phase of training. Clearly the system can be made finer by introducing a prediction model every semester or even every trimester or it can be extended to other faculty or other types of students.

Acknowledgements

Research group is composed by the authors of the contribution that was edited in the following order: Mauro Mezzini (sections 3-4 excluding subsection 3.2), Gianmarco Bonavolontà (sections 2 and subsection 3.2), Francesco Agrusti (section 1).

We wish to thank the Statistical Office of R3U for the support they give us by producing all the data that were used in this research.

References

Agrusti, F., Bonavolontà, G., & Mezzini, M. (2019). University Dropout Prediction through Educational Data Mining Techniques: A Systematic Review. Journal of E-Learning and Knowledge Society, 15(3), 161-182. https://doi.org/10.20368/1971-8829/1135017

Alban, M., Mauricio, D., Mauricio, D., and National University of San Marcos, Artificial Intelligence Group, Per; (2019). Predicting University Dropout trough Data Mining: A systematic Literature. *Indian Journal of Science and Technology*, 12(4):1–12.

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185.

ANVUR (2018). *Rapporto Biennale 2018 ANVUR Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca*.

Araque, F., Roldan, C., and Salguero, A. (2009). Factors influencing university drop out rates. *Computers & Education*, 53(3):563–574.

Bala, M. and Ojha, D. (2012). Study of applications of data mining techniques in education. *International Journal of Research in Science and Technology*, 1(4):1– 10.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.

Bandura, A., Barbaranelli, C., Caprara, G. V., and Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child development*, 72(1):187–206.

Bean, J. P. (1988). *Leaving college: Rethinking the causes and cures of student attrition*. Taylor & Francis.

Bengio, Y. and LeCun, Y., editors (2015). *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Bentler, P. M. and Speckart, G. (1979). Models of attitudebehavior relations. *Psychological review*, 86(5):452.

Bernardo Gutirrez, A. B., Esteban Garca, M., Gonzlez Garca, J. A., Nez Prez, J. C., and Dobarro Gonzlez, A. (2017). University dropout: Definition, features and prevention measures. *Factors affecting academic performance*.

Braxton, J. M., Shaw Sullivan, A. V., and Johnson, R. M. (1997). Appraising tinto's theory of college student departure. *Higher Education - New York-Agathon Press Incorporated*, 12:107–164.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231.

Bridgeman, B., McCamley-Jenkins, L., and Ervin, N. (2000). Predictions of freshman grade-point average from the revised and recentered sat R i: Reasoning test. *ETS Research Report Series*, 2000(1):i–16.

Burgalassi, M., Biasi, V., Capobianco, R., and Moretti, G. (2016). Il fenomeno dell'abbandono universitario precoce. Uno studio di caso sui corsi di laurea del Dipartimento di Scienze della Formazione dell'Università Roma Tre. *Giornale Italiano di Ricerca Didattica/Italian Journal of Educational Research*, 17:131–152.

Cabrera, A. F., Castaneda, M. B., Nora, A., and Hengstler, D. (1992). The convergence between two theories of college persistence. *The journal of higher education*, 63(2):143–164.

Camara, W. J. and Echternacht, G. (2000). The sat [r] i and high school grades: Utility in predicting success in college. research notes.

Carbone, V. and Piras, G. (1998). Palomar project: predicting school renouncing dropouts, using the artificial neural networks as a support for educational policy decisions. *Substance use & misuse*, 33(3):717–750.

Chow, C. K. and Liu, C. N. (1968). Approximating Discrete Probability Distributions With Dependence Trees. *IEEE Transactions on Information Theory*, IT-14:462–467.

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual review of psychology*, 51(1):171–200.

Cox, B. E. and Orehovec, E. (2007). Faculty-student interaction outside the classroom: A typology from a residential college. *The review of higher education*, 30(4):343–362.

Cukusic, M., Garaca, Z., and Jadric, M. (2014). Online self-assessment and students' success in higher education institutions. *Computers & Education*, 72:100–109.

DesJardins, S. L., Ahlburg, D. A., and McCall, B. P. (1999). An event history model of student departure. *Economics of education review*, 18(3):375–390.

Duque, L. C., Duque, J. C., and Suriach, J. (2013). Learning outcomes and dropout intentions: an analytical model for Spanish universities. *Educational studies*, 39(3):261–284.

Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29(2):131–163.

Georg, W. (2009). Individual and institutional factors in the tendency to drop out of higher education: a multilevel analysis using data from the Konstanz Student Survey. *Studies in Higher Education*, 34(6):647–661.

Ghignoni, E. (2017). Family background and university dropouts during the crisis: the case of italy. *Higher Education*, 73(1):127–151.

Gifford, D. D., BricenoPerriott, J., and Mianzo, F. (2006). Locus of control: Academic achievement and retention in a sample of university first-year students. *Journal of College Admission*, 191:18–25.

Gutierrez, A. B. B., Menendez, R. C., Rodrıguez-Muniz, L. J., Perez, J. C. N., Herrero, E. T., and Garcıa, M. E. (2015). Prediccion del abandono universitario: variables explicativas y medidas de prevencion. *Revista Fuentes*, (16):63–84.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141.

Hu, S. and Kuh, G. D. (2002). Being (dis) engaged in educationally purposeful activities: The influences of student and institutional characteristics. *Research in Higher Education*, 43(5):555–575.

Ishitani, T. T. (2006). Studying attrition and degree completion behavior among first-generation college students in the united states. *The Journal of Higher Education*, 77(5):861–885.

Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., and Rose, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(4):333–353.

Krause, K. L. (2005). Serious thoughts about dropping out in first year : trends, patterns and implications for higher education. *Studies in Learning, Evaluation, Innovation and Development*, 2(3):55–68.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc.

Kuncel, N. R. and Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, 315(5815):1080–1081.

Kuncel, N. R., Crede, M., and Thomas, L. L. (2007). A meta-analysis of the predictive validity of the graduate management admission test (gmat) and undergraduate grade point average (ugpa) for graduate student academic performance. *Academy of Management Learning & Education*, 6(1):51–68.

Kuncel, N. R., Hezlett, S. A., and Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of personality and social psychology*, 86(1):148.

Larsen, M. R., Sommersel, H. B., and Larsen, M. S. (2013). *Evidence on dropout phenomena at universities*. Danish Clearinghouse for educational research Copenhagen.

LeCun, Y., Bengio, Y., and Hinton, G. E. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lohfink, M. M. and Paulsen, M. B. (2005). Comparing the determinants of persistence for first-generation and continuing-generation students. *Journal of College Student Development*, 46(4):409–428.

Maas, A. L., Hannun, A. Y., and Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*.

Malvestuto, F. M., Mezzini, M., and Moscarini, M. (2011). Computing simple-path convex hulls in hypergraphs. *Inf. Process. Lett.*, 111(5):231–234.

Marshall, M. A. and Brown, J. D. (2004). Expectations and realizations: The role of expectancies in achievement settings. *Motivation and Emotion*, 28(4):347– 361.

Martinho, V. R. D. C., Nunes, C., and Minussi, C. R. (2013). An intelligent system for prediction of school dropout risk group in higher education classroom based on artificial neural networks. In *2013 IEEE 25th International Conference on Tools with Artificial Intelligence*, pages 159–166. IEEE.

Mezzini, M. (2007). Finding a nonempty algebraic subset of an edge set in linear time. *J. Graph Algorithms Appl.*, 11(1):239– 257.

Mezzini, M. (2010). On the complexity of finding chordless paths in bipartite graphs and some interval operators in graphs and hypergraphs. *Theor. Comput. Sci.*, 411(79):1212–1220.

Mezzini, M. (2011). Fast minimal triangulation algorithm using minimum degree criterion. *Theor. Comput. Sci.*, 412(29):3775– 3787.

Mezzini, M. (2012). Fully dynamic algorithm for chordal graphs with O(1) querytime and o(n^2) update-time. *Theor. Comput. Sci.*, 445:82–92.

Mezzini, M. (2016). On the geodetic iteration number of the contour of a graph. *Discrete Applied Mathematics*, 206:211–214.

Mezzini, M. (2018). Polynomial time algorithm for computing a minimum geodetic set in outerplanar graphs. *Theor. Comput. Sci.*, 745:63–74.

Mezzini, M. and Moscarini, M. (2010). Simple algorithms for minimal triangulation of a graph and backward selection of a decomposable markov network. *Theor. Comput. Sci.*, 411(7-9):958–966.

Mezzini, M. and Moscarini, M. (2015). On the geodeticity of the contour of a graph. *Discrete Applied Mathematics*, 181:209–220.

Mezzini, M. and Moscarini, M. (2016). The contour of a bridged graph is geodetic. *Discrete Applied Mathematics*, 204:213–215.

Mezzini, M., Bonavolontà, G., Agrusti F. (2019) Predicting University Dropout by using Convolutional Neural Networks, INTED2019 Proceedings, pp. 9155-9163.

Mitchell, T. M. (1997). *Machine learning*. McGraw hill.

Moretti, G., Burgalassi, M., and Giuliani, A. (2017). Enhance Students' Engagement To Counter Dropping-Out: A Research At Roma Tre University. pages 305–313, Valencia, Spain.

Mustafa, M. N., Chowdhury, L., and Kamal, M. S. (2012). Students dropout prediction for intelligent system from tertiary level in developing country. In *2012 International Conference on Informatics, Electronics & Vision (ICIEV)*, pages 113–118. IEEE.

Nagy, M. and Molontay, R. (2018). Predicting dropout in higher education based on secondary school performance. In *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*, pages 000389–000394. IEEE.

Pascarella, E. T. and Terenzini, P. T. (1980). Predicting freshman persistence and voluntary dropout decisions from a theoretical model. *The journal of higher education*, 51(1):60–75.

Pascarella, E. T. and Terenzini, P. T. (2005). *How College Affects Students: A Third Decade of Research. Volume 2.* ERIC.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Pereira, R. T., Romero, A. C., and Toledo, J. J. (2013). Extraction student dropout patterns with data mining techniques in undergraduate programs. In *KDIR/KMIS*, pages 136–142.

Perfetto, G. (2002). Predicting academic success in the admissions process: Placing an empirical approach in a larger process. *College Board Review*, 196:30–35.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In *Handbook of self-regulation*, pages 451–502. Elsevier.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1):145 – 151.

Quinlan, J. R. (2014). *C4. 5: programs for machine learning*. Elsevier.

Rosenbaum, J. (2004). Its time to tell the kids: If you dont do well in high school, you wont do well in college (or on the job). *American Educator*, 28:8–42.

Rovira, S., Puertas, E., and Igual, L. (2017). Data-driven system to predict academic grades and dropout. *PLoS one*, 12(2):e0171207.

Santelices, M. V., Cataln, X., Kruger, D., and Horn, C. (2016). Determinants of persistence and the role of financial aid: lessons from Chile. *Higher Education*, 71(3):323– 342.

Scutari, M. (2010). Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software*, 35(3):1–22.

Siri, A. (2015). Predicting students' dropout at university using artificial neural networks. *Italian Journal of Sociology of Education*, 7(2).

Sivakumar, S., Venkataraman, S., and Selvaraj, R. (2016). Predictive modeling of student dropout indicators in educational data mining using improved decision tree. *Indian Journal of Science and Technology*, 9(4):1–5.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT press, 2nd edition.

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). Inceptionv4, inception-resnet and the impact of residual connections on learning. In Singh, S. P. and Markovitch, S., editors, *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 4278–4284. AAAI Press.

Tan, P.-N., Steinbach, M., Kumar, V. (2005). *Introduction to Data Mining*. Addison Wesley. ISBN: 0321321367.

Tanucci, G. (2006). Orientamento e carriera universitaria. In Fasanella A. e Tanucci G., a cura di, *Orientamento e carriera universitaria. Ingressi ed abbandoni in cinque Facoltà dell'Università di Roma La Sapienza nel nuovo assetto didattico.* Milano: FrancoAngeli.

Ting, S.-M. R. and Robinson, T. L. (1998). First-year academic success: A prediction combining cognitive and psychosocial variables for caucasian and african american students. *Journal of college student development*.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, 45(1):89–125.

Tinto, V. (2010). From theory to action: Exploring the institutional conditions for student retention. In *Higher education: Handbook of theory and research*, pages 51–89. Springer.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, 92(4):548.

Willcoxson, L. (2010). Factors affecting intention to leave in the first, second and third year of university studies: A semesterby-semester investigation. *Higher Education Research & Development*, 29(6):623–639.