Utilizzo delle reti neurali convolutive per la predizione dell'abbandono universitario. Una ricerca quantitativa sui corsi di laurea del Dipartimento di Scienze della Formazione dell'Università «Roma Tre»

Use of convolutive neural networks to predict university dropout. A quantitative research on the degree courses of the Department of Education of "Roma Tre" University

Mauro Mezzini
Università degli Studi Roma TRE (IT), Dipartimento di Scienze della Formazione, mauro.mezzini@uniroma3.it
Gianmarco Bonavolontà
Università degli Studi Roma TRE (IT), Dipartimento di Scienze della Formazione
Francesco Agrusti
Università degli Studi Roma TRE (IT), Dipartimento di Scienze della Formazione

abstract

The level of dropout in the scenery of European education is one of the major issues to be faced in a near future. In 2017, an average of 10.6% of young people (aged 18-24) in the EU-28 were early leavers from education and training according to Eurostat's statistics.
The main aim of this research is to predict, as early as possible, which student will dropout in a Higher Education context. The administrative data of approximately 6000 students enrolled from 2009 in the Education Department at Rome Tre University had been used to train the Convolutional Neural Networks (CNN). Then, the trained network provides a probabilistic model that indicates, for each student, the probability of dropping out. We used several types of CNNs, and their variants, in order to build the most accurate model for the dropout prediction. The accuracy of the obtained models ranged from 67.1% for the students at the beginning of the first year up to 88.7% for the students at the end of the second year of their academic career.

Keywords: University Dropout; Convolutional Neural Networks; Artificial Intelligence.

Il gruppo di ricerca è formato dagli autori del contributo che è stato redatto nel seguente ordine: Mauro Mezzini (§§ 4-5-6), Gianmarco Bonavolontà (§§ 2-3), Francesco Agrusti (§§ 1).

Italian Journal of Special Education for Inclusion | © Pensa MultiMedia Editore srl | ISSN 2282–6041 (on line) | DOI: 10.7346/sipes-02-2019-32

1. Introduzione

Il fenomeno noto come "dispersione universitaria" comprende diverse forme di insuccesso accademico, riconducibili a quattro categorie principali (Fasanella et. al., 2010): l'irregolarità generica nell'acquisizione dei crediti formativi, il prolungamento della permanenza nell'università (c.d. *fuoricorsismo*), la mancanza di linearità della carriera (ad es. passaggi di corso) ed infine l'abbandono vero e proprio del percorso di apprendimento che porta quindi all'uscita dal sistema universitario senza il conseguimento del titolo.

Ci sono diverse variabili che influenzano la decisione degli studenti di lasciare gli studi a livello universitario. Questo fenomeno è noto in lingua anglosassone come "drop-out" ed è stato definito da Larsen e da altri ricercatori come "il ritiro da un corso di laurea prima che sia stato completato" (Søgaard Larsen & Dansk Clearinghouse, 2013, p. 18). In questa definizione è incluso anche il ritiro dai singoli corsi di studio ma non l'abbandono per gravidanza, malattia, ecc. ossia per tutte quelle cause che si possono ascrivere a motivazioni ben precise e di durata temporanea. Il fenomeno dell'abbandono ha diversi effetti negativi: ha conseguenze a livello personale, familiare e, da un punto di vista sistematico, questi bassi tassi di completamento del percorso universitario potrebbero portare ad un collo di bottiglia delle competenze della cittadinanza intera che può avere conseguenze sul piano economico e sociale, diminuendo la competitività, l'innovazione e la produttività di una intera nazione.

Nello studio comparativo sull'abbandono dell'istruzione superiore in Europa condotto da Vossensteyn e altri ricercatori (2015) è stato riscontrato che il successo degli studi è considerato come un fattore cruciale per il successo personale in 28 dei 35 paesi partecipanti. Un riconoscimento precoce del fenomeno dell'abbandono è il prerequisito fondamentale per ridurre i tassi dell'abbandono stesso: diversi studi evidenziano l'importanza di monitorare le caratteristiche individuali e sociali degli studenti in quanto queste hanno un forte impatto sulle probabilità di successo degli studenti nell'istruzione superiore. Un obiettivo cruciale della strategia Europa 2020 è difatti quello di ridurre l'abbandono degli studi universitari cercando di ottenere almeno che il 40% dei cittadini di 30-34 anni completino il percorso di istruzione superiore (Vossensteyn et al., 2015).

Come riportato in letteratura, gli studenti in generale lasciano durante il loro primo anno di università (Larsen et al., 2013), subito dopo la scuola secondaria superiore: in questo periodo, devono sviluppare il loro senso di responsabilità e autoregolamentazione (Moè & De Beni, 2000). Le competenze e le disposizioni individuali sono indagate in diversi modelli psicologici e pedagogici in relazione al fenomeno dell'abbandono precoce in termini di caratteristiche della personalità (Pincus, 1980). Numerose ricerche hanno esplorato l'impatto dello status economico e sociale degli studenti (es. razza o reddito) e dei servizi di carattere organizzativo forniti agli studenti dall'università (es. rapporto facoltà-studenti) sul tasso di abbandono (Pincus, 1980; Stampen & Cabrera, 1988).

Da decenni uno dei modelli più utilizzati e discussi è il modello "student integration" di Tinto che sottolinea l'importanza dell'integrazione accademica e sociale degli studenti nella previsione del fenomeno dell'abbandono scolastico (Tinto, 1975, 1987, 2010). Uno degli altri modelli principali è quello proposto da Bean (1968), il modello "student attrition", basato sull'atteggiamento-comporta-

mento dello studente che misura i fattori individuali e quelli istituzionali e ne valuta le interazioni ai fini della previsione dell'abbandono universitario. Un altro modello interessante di integrazione studente/istituzione è il modello di Pascarella (1980), che pone l'accento sulla crucialità per il successo degli studenti l'avere contatti informali con i docenti. In altre parole, in questo modello le caratteristiche di background interagiscono con i fattori istituzionali influendo sulla soddisfazione dello studente nei confronti dell'università. Numerosi studi hanno dimostrato gli effetti positivi dell'interazione studente-università sulla persistenza (Astin, 1993; Cox & Orehovec, 2007; Pascarella & Terenzini, 1991, 2005; Braxton, Sullivan & Johnson, 1997; Milem & Berger, 1997). L'*Event history modeling* è un altro modello molto discusso in letteratura: proposto da Des Jardins, Albourg e Mccallan (1999), questo modello tiene conto del ruolo della successione di diversi eventi nelle diverse fasi della carriera formativa dello studente, cambiando l'importanza dei fattori di anno in anno, in base al periodo temporale.

In tutti questi modelli, la relazione tra studenti e istituzioni è di cruciale importanza per ridurre i tassi di abbandono scolastico (Cabrera et al., 1992) e sono state identificate diverse variabili per migliorare la ritenzione degli studenti (Søgaard Larsen & Dansk Clearinghouse, 2013; Siri, 2014).

Da numerose ricerche statunitensi (Hu, 2002; Kuncel et al., 2004; Bridgeman, McCamley-Jenkins, & Ervin, 2000; Kuncel & Hezlett, 2007; Kuncel, Credé, & Thomas, 2007), il voto di maturità si è dimostrato come il miglior predittore della *performance* del primo anno accademico (predicendo meglio dei punteggi standardizzati SAT) e più nello specifico del voto medio che lo studente ottiene al primo anno di *college* (Perfetto, 2002). Il legame però tra voto di maturità e persistenza nel sistema formativo resta un argomento più controverso: Rosenbaum (2004) asserisce che "il predittore della probabilità che uno studente si laurei più facile da usare è ancora il suo voto di maturità" (p. 2); allo stesso modo Ishitani (2006) afferma che "la posizione in graduatoria in classe alla maturità ha effetti significativi sul comportamento di *attrition* universitaria" (p. 18). Contemporaneamente però esistono altre ricerche in letteratura che ritengono indicatori insufficienti il voto di maturità e i punteggi a test standardizzati (ad esempio il SAT) per predire la persistenza all'università (Lohfink & Paulsen, 2005; Allen et al., 2008).

In Italia, a causa degli altissimi tassi di abbandono degli studenti universitari (ANVUR, 2018), sono stati condotti diversi studi specifici (Burgalassi et al., 2016; Moretti et al., 2017; Carbone & Piras, 1998) che hanno confermato la valenza del voto di maturità (e delle competenze in ingresso degli studenti più in generale) insieme ai tratti socio-demografici degli studenti (maggiormente il contesto socio-economico) come validi indicatori dell'abbandono universitario rispetto all'esito del primo anno di studi.

Molti dei modelli e degli studi condotti, sia nazionali sia internazionali, hanno presentato diverse analisi dal punto di vista psicologico, costruendo modelli psicologico-motivazionali focalizzate sull'aspettativa, sulle ragioni del coinvolgimento, sul valore personale e sulla motivazione in generale (Bandura et al., 2001; Marshall & Brown, 2004; Gifford et al., 2006; Yorke, 2002; Anderman et al. 2001; Pintrich, 2000; Le et al., 2005; Robbins et al., 2004). Tali modelli e indagini prevedono tutti una rilevazione dei dati intervistando direttamente gli studenti, tramite l'uso di strumenti (di solito questionari) appositamente somministrati. Lo studio

presentato in questo articolo, invece, si prefigge di utilizzare i soli dati disponibili in un qualsiasi ufficio statistico universitario, senza quindi, almeno in questa fase della ricerca, intervistare direttamente gli studenti. A questo proposito si è quindi deciso di procedere all'analisi di questi dati tramite l'uso dell'Intelligenza Artificiale (IA).

Oggi, l'IA viene utilizzata per sostituire le attività umane che sono ripetitive, ad esempio, nel campo di guida autonoma o per il compito di classificazione delle immagini. In queste aree l'IA compete con l'uomo con risultati abbastanza soddisfacenti e, nel caso di abbandono del sistema formativo, è estremamente improbabile che un docente esperto sia in grado di "prevedere" il successo educativo dello studente sulla base dei soli dati forniti dagli uffici amministrativi.

La recente scoperta sulle reti neurali (RN) con l'uso di architetture di Reti Neurali Convoluzionali (CNN, *Convolutional Neural Networks*) è diventata, negli ultimi anni, a dir poco dirompente negli studi che utilizzano l'IA. Impilando insieme decine o centinaia di strati neurali convoluzionali, si ottiene una struttura di rete profonda, che si è dimostrata molto efficace nel produrre modelli ad alta precisione.

Questi recenti progressi sulle reti neurali hanno dimostrato che l'IA può essere in grado di competere (o addirittura superare) con le capacità umane, nei compiti di classificazione e riconoscimento. Qui di seguito sono quindi dapprima mostrati alcuni degli studi risultati più importanti, ottenuti grazie all'IA, sulla previsione dell'abbandono universitario. Successivamente sono presentate le metriche per la valutazione di tali modelli e quindi la metodologia usata e i risultati ottenuti nella nostra ricerca. Sono quindi tratte brevemente delle conclusioni preliminari sullo studio.

2. Stato dell'arte

Molti progetti di ricerca hanno utilizzato tecniche di *data mining* per studiare il fenomeno del Dropout. Nello specifico, in questo paragrafo discuteremo di lavori che hanno indagato sull'abbandono universitario sviluppando modelli predittivi attraverso l'EDM (*Educational Data Mining*), ovvero l'utilizzo del *data mining* nel campo dell'educazione, applicando metodi computerizzati per analizzare ampie raccolte di dati (Bala & Ojha, 2012; Koedinger et al., 2015). Dall'analisi della letteratura è emerso che l'algoritmo dell'albero decisionale (DT) sia quello più utilizzato per lo sviluppo di modelli predittivi finalizzato ad individuare l'abbandono universitario. Una ricerca condotta presso l'Università di Chittagong ha esaminato la possibilità di predire il successo formativo degli studenti neoiscritti all'università utilizzando solo i dati presenti al momento dell'iscrizione. Gli algoritmi utilizzati per lo sviluppo di questi modelli sono stati il CART (Classification And Regression Tree) e CHAID (Chi-Squared Detector Automatic Interaction Detector), con l'utilizzo del *cross-validation folder* per decidere quale modello sia migliore in termini di accuratezza (Mustafa et al., 2012). Un altro progetto di ricerca finanziato dal ministero della Pubblica Istruzione colombiana ha cercato di identificare i modelli predittivi di abbandono scolastico analizzando 62 attributi appartenenti a dati socio-economici, accademici ed istituzionali. Anche in questo caso è stato implementato un albero decisionale (algoritmo J48) e per la validazione del modello si

è utilizzato il *cross-validation folder* con un risultato di accuratezza superiore al 80% (Pereira et al., 2013). Allo stesso modo, in India è stata condotta una ricerca per sviluppare un DT basato su l'algoritmo ID3 in grado di prevedere gli studenti che avessero abbandonato l'università. Lo studio si basa sull'analisi di 32 variabili su un campione di 240 studenti scelti attraverso un'indagine. Le prestazioni dei modelli sono state valutate usando l'indice di accuratezza, la precisione, il *recall* e l'*F-measure* (Sivakumar et al., 2016). Nel 2018 una ricerca ha presentato una classificazione basata sull'algoritmo DT. Lo studio analizza 5288 casi di studenti appartenenti all'università pubblica cilena (coorti di studenti appartenenti a 44 corsi universitari nei settori delle discipline umanistiche, delle arti, dell'istruzione, dell'ingegneria e della salute). Gli attributi selezionati per l'analisi sono legati alle variabili demografiche dello studente, alla sua situazione economica e ai dati sul rendimento scolastico pregresso prima della sua ammissione all'università. L'indice di accuratezza del miglior modello sviluppato è stato del 87,2% (Ramírez & Grandón, 2018).

In aggiunta al DT sono stati utilizzati altri metodi di classificazione al fine di implementare modelli per la predizione dell'abbandono universitario. Alcuni ricercatori hanno utilizzato metodologie specifiche come CRISP-DM (*Cross Industry Standard Process for Data Mining*), per prevedere alla fine del primo semestre gli studenti a rischio di abbandono. Il dataset è composto da oltre 25 mila studenti e 39 variabili per ogni studente e gli algoritmi utilizzati sono: DT, reti neurali artificiali (ANN) e modello logit (LR). I risultati evidenziano un'accuratezza pari al 81,2% per il modello sviluppato con ANN (Delen, 2011). Similarmente, una ricerca condotta presso l'Università di Genova, ha impiegato le ANN per rilevare gli studenti a rischio di abbandono. Lo studio fa riferimento ad una popolazione che è composta da 810 studenti iscritti per la prima volta in un corso di laurea in medicina nell'anno accademico 2008-2009 e i dati provengono da fonti amministrative, da un sondaggio ad hoc e da interviste telefoniche (Siri, 2015). Altro esempio è il lavoro svolto presso il College of Technology nel Mato Grosso. La ricerca presenta un modello sviluppato con la rete neurale Fuzzy-ARTMAP utilizzando solo i dati di immatricolazione raccolti per un periodo di sette anni dal 2004 al 2011. I risultati mostrano una percentuale di accuratezza superiore all'85% (Martinho et al., 2013). In Brasile presso Universidade Federal do Rio de Janeiro, un progetto di ricerca ha confrontato diversi algoritmi (DT, SimpleCart, Support Vector Machine, Naïve Bayes e ANN) analizzando dati provenienti da 14.000 studenti (Manhães et al., 2014). Analogamente, presso l'Università di Tecnologia ed Economia di Budapest, utilizzando i dati di 15.285 studenti universitari riguardo la loro istruzione secondaria e universitaria, sono stati impiegati e valutati 6 tipi di algoritmi per identificare gli studenti a rischio dropout. L'accuratezza, il *recall*, la precisione e la curva di ROC sono le metriche utilizzate per la valutazione ed i risultati hanno evidenziato come miglior modello quello sviluppato dall'algoritmo di *Deep Learning* con un tasso di accuratezza del 73,5% (Nagy & Molontay, 2018). Una ricerca simile ha impiegato cinque algoritmi di classificazione (LR, Gaussian Naive Bayes, SVM, Random Forest e Adaptive Boosting) analizzando 4432 dati degli studenti dei corsi di laurea in Giurisprudenza, Informatica e Matematica dell'università di Barcellona negli anni 2009 e 2014. La ricerca ha rilevato che tutti gli algoritmi di *machine learning* raggiungevano un'accuratezza intorno al 90% (Rovira et al., 2017). l'Instituto Tecnológico de Costa Rica ha implementato un modello derivato

dagli algoritmi di Random Forest, Support Vector Machine, ANN e LR. Il campione di riferimento è composto da 16.807 studenti iscritti tra gli anni 2011 e 2016 ed il modello migliore è quello risultante dall'algoritmo di Random Forest (Solis et al., 2018).

Gli studi sopra citati evidenziano un uso eterogeneo di *dataset*, algoritmi, metriche e metodologie di performance. Pertanto, diviene improbabile definire con certezza quale modello sia migliore dell'altro e tuttavia le ricerche confermano l'efficacia dell'approccio dell'EDM finalizzato allo studio del dropout universitario. La principale differenza che caratterizza questo lavoro da quelli presenti in letteratura è data dall'introduzione delle reti neurali convolutive per analizzare dati appartenenti al campo educativo.

3. Metriche per la valutazione dei modelli

La valutazione delle prestazioni del modello di classificazione si basa principalmente sul conteggio dei casi classificati correttamente ed incorrettamente. Questi dati sono rappresentati attraverso la matrice di confusione, ovvero la tabulazione delle classi reali e delle classi predette, dove in riga sono riportati i valori reali o effettivi della classe mentre in colonna sono riportati le previsioni fornite dal modello.

		Classe Predetta	
		-1	+1
Classe Reale	-1	TN	FP
	+1	FN	TP

Nel caso di un problema di classificazione binaria, la matrice di confusione è composta da quattro diversi elementi: veri negativi (*true negative*, TN), falsi positivi (*false positive*, FP), falsi negativi (*false negative*, FN) e veri positivi (*true positive*, TP). entrando nello specifico, TN rappresenta il numero di casi il cui valore della classe reale è negativo (-1) ed il modello predice una classe negativa (-1); FP rappresenta il numero di casi in cui il valore della classe reale è negativo (-1) ma il modello predice erroneamente la classe positiva (+1); FN rappresenta il numero di casi in cui il valore della classe reale è positivo (+1) mentre il modello predice in modo errato (-1); TP rappresenta il numero di casi in cui la classe reale e quella predetta sono entrambe positive (+1).

Utilizzare metriche che derivano dalla matrice di confusione rende più conveniente la valutazione delle prestazioni di diversi modelli di classificazione. La prima metrica da prendere in considerazione è l'accuratezza (ACC). L'accuratezza è importante perché misura la capacità del modello di classificazione di fornire previsioni attendibili su nuovi dati ed è data dalla somma delle previsioni corrette divisa per il numero totale delle previsioni:

$$ACC = \frac{TP + TN}{FP + FN + TP + TN}$$

Di conseguenza, l'accuratezza descrive la percentuale di casi correttamente classificati indipendentemente dalla produzione di falsi positivi o falsi negativi. Da ciò ne deriva che l'accuratezza è una misura che a volte potrebbe essere fuorviante e pertanto sono utilizzate, in aggiunta a questa, altre metriche di valutazione come la precisione (PRE) e la *recall* (REC). Queste metriche sono correlate ai rapporti di veri positivi e veri negativi. Nello specifico, la precisione è la proporzione fra i veri positivi e tutti i valori classificati come positivi:

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

Più alto è il valore della precisione e minore è il numero dei FP. Siamo nella situazione in cui è prioritario classificare correttamente i veri positivi, anche al costo di creare un elevato numero di falsi positivi.

L'indice di *recall* è la proporzione fra i veri positivi e tutti i valori che sono effettivamente positivi:

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

La *recall* descrive l'efficacia del modello nel riconoscere la proprietà osservata e pertanto siamo nel caso in cui si vuole massimizzare i TP senza però far crescere troppo i FP. Maggiore è il valore della *recall*, più sensibile è il modello nel predirre i TP. Esiste un'altra metrica chiamata **F_1-measure** che riassume l'indice di precisione e di *recall* e rappresenta la media armonica tra i due indici:

$$F_1\text{measure} = \frac{2 * sensibility * precision}{sensibility + precision}$$

Un valore elevato dell'**F_1-measure** indica che sia la *recall* sia la precisione hanno un valore ragionevolmente alto e pertanto l'indice **F_1-measure** è utile quando PRE e REC sono ugualmente importanti.

4. Metodologia

Uno dei problemi più importanti nel campo dell'IA è il problema della *classificazione* (LeCun et al., 2015). In questo problema si ha un oggetto, che può essere un'immagine, un suono o una frase e si vuole associare a questo oggetto una classe presa all'interno di un insieme finito $\boldsymbol{K}$ di classi. Per esempio, il problema della classificazione dell'immagini consiste nell'associare a ciascuna immagine una certa classe in accordo con una prestabilita interpretazione. In questo caso, una naturale interpretazione sarebbe quella di associare a ciascuna immagine il soggetto in essa contenuto. Messo in termini più matematici, se si rappresenta ogni oggetto mediante un vettore $\boldsymbol{n}$– dimensionale di numeri reali $\boldsymbol{x}\boldsymbol{\epsilon}\mathbb{R}^n$, la soluzione del problema della classificazione consiste nel trovare una funzione equivalente alla funzione $f:\mathbb{R}^n\rightarrow\boldsymbol{K}$ che associa ad ogni oggetto $\boldsymbol{x}$ la sua classe vera.

Una rete neurale (RN) può essere vista infatti come una funzione φ che prende in input un vettore $\boldsymbol{n}$– dimensionale $\boldsymbol{x}$ e produce un valore, chiamato la predizione di $\boldsymbol{x}$. La predizione è corretta quando $\varphi(x)=f(x)$ and incorretta altrimenti. Contrariamente al paradigma classico della programmazione, dove il programmatore per progettare un algoritmo deve avere una profonda e completa conoscenza del problema di interesse come ad esempio in (Malvestuto, Mezzini, & Moscarini, 2011; Mezzini, 2010, 2011, 2012, 2016, 2018; Mezzini & Moscarini, 2015, 2016), per implementare una RN il programmatore può anche essere del tutto ignaro del meccanismo o della semantica di classificazione.

Per fare sì che una RN produca corrette predizioni questa deve essere sottoposta ad un processo di *addestramento* (training). Questo consiste nel fornire alla RN un insieme di oggetti chiamato *insieme di addestramento* (*training set*) e indicato come $A=(x_i,f(x_i)),i=1,\ldots,N$ dove N è il numero degli oggetti nell'insieme di addestramento. La classe $f(x_i)$, di ogni oggetto x_i nell'insieme di addestramento, è nota a priori. Per ogni oggetto nell'insieme di addestramento, il valore $f(x_i)$ viene confrontato con la predizione $\varphi(x_i)$ della RN. Se il valore $\varphi(x_i)$ della predizione è differente dalla sua classe $f(x_i)$, la RN viene modificata al fine di minimizzare l'errore. Questo processo viene ripetuto centinaia di volte fino a che viene raggiunto un livello prefissato di errore oppure quando il livello di errore non migliora. Questo processo viene chiamato *apprendimento supervisionato* (*supervised learning*) ed è simile al processo di apprendimento che viene impiegato o negli esseri umani o negli animali.

Tra i differenti tipi di RN le RNC hanno acquistato molta popolarità negli ultimi anni grazie agli ottimi risultati ottenuti nella classificazione delle immagini (Krizhevsky et at., 2012). Krizhevsky et al. hanno addestrato un modello convoluzionale profondo utilizzando un insieme di addestramento composto da 1.2 milioni di immagini della *challenge Imagenet* contenente 1000 differenti classi e migliorando l'accuratezza della predizione di quasi il 50% in più rispetto ai precedenti sistemi di classificazione. Da allora ad oggi molte ricerche sono state realizzate sulle RNC.

Nelle RN classiche l'oggetto di input viene inviato, all'inizio, ad un insieme di n_1 neuroni, chiamato il *primo livello*. Ogni neurone nel primo livello riceve in input una copia dell'oggetto $\boldsymbol{x}$, e utilizzando un vettore $\boldsymbol{n}$– dimensionale w_j, $j=1,\ldots,n_1$ di pesi, produce in output un numero reale chiamato *attivazione*. Pertanto, il pri-

mo livello fornisce in uscita un vettore n_1– dimensionale di reali. Quest'ultimo vettore può essere inviato ad un secondo livello di neuroni e così via. In questo modo possiamo accumulare una pila di livelli neuronali per ottenere una rete più complessa e potente. Il problema è che il numero di pesi della rete e quindi le risorse di memoria e computazionali richieste possono divenire molto elevate tanto più cresce il numero di livelli della rete e tanto più grande è il numero di neuroni in ciascun livello. Ad esempio, alla risoluzione di 32x32 *pixels*, un'immagine è considerata molto piccola anche per i più vecchi *smartphone*. Ciò nonostante la codifica numerica di una tale immagine consiste in un vettore 3072-dimensionale. Se nel primo livello volessimo introdurre 1000 neuroni, il numero di pesi solo per il primo livello sarebbe di 3 milioni. La differenza chiave tra le RNC e le RN è che non tutte le componenti dell'oggetto sono inviate a ciascun neurone, ma soltanto una parte di queste. In questo modo il numero di pesi per ciascun neurone diminuisce drasticamente ma allo stesso tempo ogni livello neuronale è in grado di produrre un numero molto elevato di attivazioni.

Nel nostro lavoro abbiamo utilizzato tre differenti architetture di RNC al fine di provare la loro efficienza per la costruzione del modello predittivo. Le prime due architetture, chiamate di seguito rispettivamente ResNetV2 (RNV2) e InceptionResNetV4 (IRNV4) rappresentano lo stato dell'arte nel campo delle RNC avendo realizzato i migliori risultati possibili o tra i migliori (alla data del 2017) nelle prove di classificazione di riferimento (*benchmark*). La terza architettura, chiamata DFSV1 (Mezzini, Bonavolontà & Agrusti, 2019), è stata costruita al nostro interno attraverso alcune modificazioni delle architetture ResNet e VGG.

Abbiamo raccolto dall'ufficio amministrativo dell'Università Roma Tre (R3U), un insieme di dati relativi agli studenti iscritti presso il Dipartimento di Scienze della Formazione (DSF). Gli anni di iscrizione variano dal 2009 al 2014 per un totale di 6078 studenti dei quali 649 erano ancora attivi al momento dell'analisi dell'insieme di dati (Agosto 2018), ovvero ancora non avevano concluso il loro corso di studi, mentre i rimanenti 5429 avevano chiuso il loro corso di studi o laureandosi o abbandonando oppure per altre cause, come trasferimento ad altra università ed altro che di seguito spiegheremo meglio. Chiamiamo *non attivi* tutti gli studenti che hanno chiuso il loro corso degli studi in un modo o nell'altro. Una regola amministrativa di R3U stabilisce un limite di tempo, fissato in 9 anni, per il completamento degli studi, il che significa che uno studente può rimanere iscritto per al più 9 anni. Se uno studente non riesce a laurearsi per tale periodo il suo corso di studi viene considerato chiuso. Questi casi sono stati considerati come fossero casi di abbandono universitario.

In generale ogni studente non attivo è classificato in due possibili modi: *nonAbbandona*, *Abbandona*. Sono stati esclusi dalle due classi quelli che hanno cambiato corso di laurea o cambiato ateneo. Il numero di questi ultimi è pari a 118 studenti. Nella classe *nonAbbandona* sono conteggiati gli studenti che si sono laureati. Il numero di quelli che si sono laureati è pari a 2833 mentre quello di chi abbandona è pari a 2478.

L'ufficio amministrativo di R3U ci ha fornito molti (tra tutti quelli disponibili) degli attributi amministrativi di ciascun studente. Nella Tabella 1 è riportata la lista dei campi amministrativi utilizzati.

A causa di problemi di privacy, pur essendo disponibili, alcuni attributi sono stati censurati o non ci sono stati forniti. Infatti, non è stato possibile conoscere

il comune di nascita né il comune di residenza degli studenti. Altri attributi, come lo stato occupazionale (se lo studente è o meno occupato lavorativamente) non sono stati raccolti in modo accurato dai sistemi informativi, poiché abbiamo trovato solo pochi studenti con stato occupazionale di lavoratore quando è ben noto che un considerevole numero di studenti del DSF già lavorano come educatori nelle scuole dell'infanzia o in quelle elementari.

Lista 1	**Lista 2**
Anno di iscrizione (coorte)	Anno accademico
Anno di nascita	Codice del corso di laurea
Genere	Nome del corso di laurea
Continente di nascita	Anno di corso
Tipo di scuola superiore	Indicatore situazione economica
Voto di maturità	Studente lavoratore
Massimo voto di maturità	Esenzione dalle tasse
Anno di maturità	Tipo di esenzione dalle tasse
Trasferito da altra universita	Handicap
CFU da altra università	Part time
Facoltà	Part time CFU
	Tipo di rinnovo dell'iscrizione

Tabella 1. Lista degli attributi amministrativi

Lista 3
Insegnamento
Voto
Voto massimo
CFU
Nome classe
Anno accademico esame
Tipo di convalida
Codice tipo ric
Pregresso
Ateneo straniero
Sigla
Sovranumerari
Escludi da media
Escludi da carriera
Data esame

Tabella 2. Lista degli attributi relativi alle carriere

Si noti che, per un dato studente, i valori degli attributi della Lista 2 nella Tabella 1 possono variare di anno in anno mentre i valori degli attributi nella Lista 1 non cambiano durante l'intera carriera accademica dello studente. Gli attributi della Lista 3 nella Tabella 2 sono invece relativi alle carriere universitarie di ciascuno studente ovvero rappresentano attributi relativi a ciascuna prova d'esame o di tirocinio.

Al fine di costruire l'insieme di addestramento si è reso necessario associare a ciascuno studente una sua rappresentazione numerica. Pertanto, tutti i domini degli attributi presenti nell'insieme dei dati sono stati convertiti, attraverso una arbitraria funzione biiettiva, in un dominio intero non negativo. Ad esempio, il dominio dell'attributo *genere* che è composto dalle due parole {"*maschio*", "*femmina*"}, è stato convertito nel dominio {0,1} dove 0 corrisponde a "*maschio*" e 1 a "*femmina*".

Abbiamo poi creato una tabella *studente* il cui schema S (cioè la lista dei campi) contiene tutti i campi della Lista 1 di Tabella 1. Per ogni campo della lista Lista 2, abbiamo aggiunto ad S, 5 campi, denotati come f_y dove $y=0,..,4$, ovvero uno per ciascuno dei primi 4 anni di iscrizione universitaria. Se uno studente, per qualsiasi motivo, finisce la sua carriera nell'anno z, allora f_y è impostato al valore δ per ogni $z<y\leq 4$. Il valore di δ, che è stato scelto essere arbitrariamente pari a -1, può anche essere visto come il valore *NULLO* e non appare nel dominio di ogni altro campo dello schema *S*. Inoltre, per ogni campo della Lista 3 in Tabella 2 e per ogni anno di iscrizione $y\in\{1,..,4\}$ si è fissato un numero m_y che rappresenta il numero massimo di esami che uno studente può dare nell'anno di iscrizione y. Quindi per ogni campo della Lista 3 e per ogni anno y e per ogni $0\leq z\leq m_y$ si è aggiunto un campo $g_{y,z}$. Se uno studente, nell'anno di iscrizione $y>0$ sostiene non più di *j* esami, allora $g_{y,z}$ è impostato al valore δ per $j<z\leq m_y$. Complessivamente la tabella *studente* possiede 897 campi o colonne.

Per effettuare tutti i test abbiamo scelto una permutazione casuale degli studenti non attivi. Poi abbiamo diviso il numero degli studenti non attivi in una suddivisione $\Pi=\{P_0,P_1...,P_{11}\}$ di 12 partizioni mutuamente disgiunte ciascuna delle quali contenente 450 studenti (tranne l'ultima che contiene 479 studenti). Per ogni $0\leq i\leq 11$, la partizione P_i è stata utilizzata per rappresentare l'insieme di validazione V_i, la partizione $P_{(i+1) \bmod 12}$ è stata utilizzata per rappresentare l'insieme di test T_i e il restante per l'insieme di addestramento A_i. Da ciascuna partizione sono stati successivamente eliminati quegli studenti che ne si sono laureati ne hanno abbandonato gli studi. Spieghiamo di seguito l'utilizzo di questi tre insiemi.

Se dobbiamo predire l'abbandono universitario di uno studente al momento dell'iscrizione abbiamo a disposizione solo i suoi dati anagrafici fino all'anno 0. In generale per uno studente che inizia l'anno di iscrizione conosciamo i dati fino all'anno X. Pertanto abbiamo creato 5 tabelle denotate come *studente*_A_y, proiettando in queste tutti i campi della tabella studente relativi alla Lista 1, tutti i campi del tipo f_y e tutti i campi del tipo $g_{y,z}$ dove $y\leq x$ e scartando tutti gli altri campi. Ad esempio, lo schema della tabella *studente* _A_0 contiene solo i campi relativi alla Lista 1 della Tabella 1, e per ogni campo f della Lista 2 di Tabella 1, soltanto f_0 e nessun attributo della Lista 3 è presente. Ogni tabella *studente*_A_y, $y=0,..,4$ è stata utilizzata per selezionare gli insiemi di addestramento, di validazione e di test per l'anno y. Per avere un'idea migliore dell'efficacia dei modelli

abbiamo anche approntato un insieme di 5 tabelle nelle quali sono state eliminati tutti i campi relativi alla Lista 3 per ogni anno di iscrizione. Ovvero queste tabelle contengono solo attributi di tipo amministrativo. Abbiamo denotato queste tabelle come *studente* $\_B_y$.

Abbiamo quindi effettuato l'addestramento di 3 modelli basati su architetture RNC di cui abbiamo fatto menzione sopra per ogni anno fino all'anno 3 di iscrizione. Infatti, dopo il terzo anno di iscrizione il numero degli studenti che abbandona comincia a diventare una frazione degli studenti dell'insieme di addestramento. Inoltre, l'insieme di validazione si restringe tanto più quanto si va in avanti con gli anni di iscrizione rendendo così le statistiche meno significative. Abbiamo effettuato l'addestramento di ciascun modello fino a 100 *epoche* dove per epoca si intende che l'intero insieme di addestramento è stato elaborato o *appreso* dal modello. In altri termini tutto l'insieme di addestramento viene appreso dal modello per 100 volte e poi si arresta l'addestramento poiché, in generale, si è osservato che oltre le 100 epoche non si avverte alcun miglioramento nella accuratezza del modello. Inoltre, la migliore accuratezza sull'insieme di validazione si riscontra quasi mai alla 100 esima epoca.

Complessivamente sono state elaborate più di 58500 epoche e considerando le 12 partizioni, i 4 anni di elaborazione, le 3 differenti architetture di RNC ed i 2 tipi di schema delle tabelle, abbiamo per elaborare i risultati riportati nel presente manoscritto preso in considerazione i dati di 58500 epoche. Per ciascuna epoca è stata rilevata la matrice di confusione indicata nel paragrafo 2 sia per l'insieme di validazione che per l'insieme di test, in modo tale da poter valutare l'efficacia dei modelli e ideare una strategia di selezione del modello migliore.

La metodologia di selezione del modello migliore prevede di utilizzare un qualche indice di prestazione ottenuto dalla matrice di confusione ricavata dall'insieme di validazione. Bisogna considerare che la grandezza dell'insieme di validazione dipende anche dall'anno di iscrizione. All'atto dell'iscrizione la dimensione dell'insieme e di validazione è di poco minore a 450 studenti. All'aumentare dell'anno di iscrizione tale numero diminuisce poiché una parte degli studenti considerati o si laurea o abbandona prima dell'anno considerato.

Anno	Architett.	VALIDAZIONE				TEST			
		TP	TN	FP	FN	TP	TN	FP	FN
0	RNV2	1799	1578	1218	644	1759	1521	1275	684
0	IRNV4	2066	1150	1646	377	2022	1079	1717	421
0	DFSV1	2009	1114	1682	434	1953	1078	1718	490
1	RNV2	656	2542	254	324	621	2501	295	359
1	IRNV4	654	2531	265	326	627	2501	295	353
1	DFSV1	657	2511	285	323	626	2486	310	354
2	RNV2	354	2639	137	181	335	2621	155	200
2	IRNV4	365	2619	157	170	327	2598	178	208
2	DFSV1	322	2661	115	213	306	2661	115	229
3	RNV2	214	1151	89	90	199	1137	103	105
3	IRNV4	234	1159	81	70	202	1142	98	102
3	DFSV1	220	1159	81	84	211	1149	91	93

Tabella 3. Matrice di confusione ricavata dalla somma delle 12 partizioni per la Tabella Studente_A

Anno	Architett.	VALIDAZIONE				TEST			
		TP	TN	FP	FN	TP	TN	FP	FN
0	RNV2	1781	1604	1192	662	1710	1542	1254	733
0	IRNV4	2041	1168	1628	402	1986	1115	1681	457
0	DFSV1	2011	1144	1652	432	1954	1106	1690	489
1	RNV2	456	2252	544	524	392	2231	565	588
1	IRNV4	678	1666	1130	302	616	1639	1157	364
1	DFSV1	627	1715	1081	353	580	1701	1095	400
2	RNV2	253	2530	246	282	195	2505	271	340
2	IRNV4	316	2388	388	219	269	2304	472	266
2	DFSV1	246	2531	245	289	190	2493	283	345
3	RNV2	145	1096	144	159	111	1082	158	193
3	IRNV4	181	1026	214	123	139	1001	239	165
3	DFSV1	146	1098	142	158	112	1082	158	192

Tabella 4. Matrice di confusione ricavata dalla somma delle 12 partizioni per la Tabella Studente_B

Anno	Architett.	VALIDAZIONE				TEST		
		Avg Accuratezza	Avg F_1	DevStd F_1	Max F_1	Avg Accuratezza	Avg F_1	DevStd F_1
0	RNV2	64.5%	65.9%	1.7%	68.3%	62.6%	64.2%	2.7%
0	IRNV4	61.4%	67.0%	1.9%	69.9%	59.2%	65.4%	1.6%
0	DFSV1	59.6%	65.4%	1.6%	67.9%	57.9%	63.8%	2.6%
1	RNV2	84.7%	69.6%	5.5%	81.3%	82.6%	65.5%	6.6%
1	IRNV4	84.3%	69.0%	5.0%	79.3%	82.8%	65.9%	5.2%
1	DFSV1	83.9%	68.4%	4.1%	75.6%	82.4%	65.2%	4.0%
2	RNV2	90.4%	68.9%	4.3%	76.9%	89.3%	65.0%	6.1%
2	IRNV4	90.1%	68.8%	3.7%	75.2%	88.3%	62.7%	3.4%
2	DFSV1	90.1%	66.1%	4.4%	74.2%	89.6%	63.7%	5.3%
3	RNV2	88.4%	70.3%	6.0%	80.9%	86.6%	66.4%	8.8%
3	IRNV4	90.3%	74.9%	6.3%	84.4%	87.1%	66.5%	8.1%
3	DFSV1	89.4%	72.4%	6.5%	83.3%	88.2%	69.3%	6.1%

Tabella 5. Aggregati di accuratezza e indice F1 sull'insieme di validazione e di test della Tabella Studente_A

Anno	Architett.	VALIDAZIONE				TEST		
		Avg Accuratezza	Avg F_1	DevStd F_1	Max F_1	Avg Accuratezza	Avg F_1	DevStd F_1
0	RNV2	64.6%	65.7%	1.5%	68.2%	62.1%	63.2%	2.4%
0	IRNV4	61.2%	66.7%	2.0%	70.5%	59.2%	65.0%	2.6%
0	DFSV1	60.2%	65.8%	1.8%	69.7%	58.4%	64.2%	3.1%
1	RNV2	71.7%	46.1%	3.8%	53.1%	69.4%	40.4%	4.1%
1	IRNV4	62.1%	48.7%	2.9%	54.3%	59.7%	44.8%	4.1%
1	DFSV1	62.0%	46.5%	2.6%	51.2%	60.3%	43.1%	5.0%
2	RNV2	84.0%	49.0%	4.0%	57.1%	81.5%	38.7%	6.0%
2	IRNV4	81.6%	51.5%	3.5%	58.8%	77.8%	41.6%	7.5%
2	DFSV1	83.8%	48.2%	4.1%	58.6%	81.0%	37.4%	4.9%
3	RNV2	80.4%	48.9%	6.9%	62.2%	77.2%	38.6%	5.8%
3	IRNV4	78.3%	52.8%	5.7%	65.1%	73.7%	38.1%	12.1%
3	DFSV1	80.7%	49.6%	6.2%	65.3%	77.4%	38.2%	7.1%

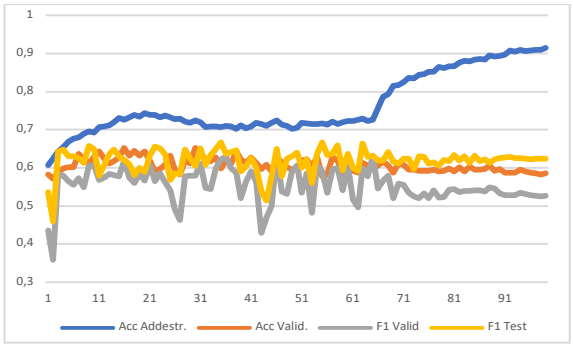
Tabella 6. Aggregati di accuratezza e indice F1 sull'insieme di validazione e di test della Tabella Studente_B

5. Risultati

L'addestramento di RNC è un processo che richiede molte risorse di tipo computazionale e quindi anche risorse temporali. Pertanto, si sono utilizzate diverse GPU per implementare tutti i test effettuati. Abbiamo utilizzato lo stato dell'arte del software per la implementazione di RNC. Tutti gli esperimenti sono stati realizzati utilizzando il software TensorFlow con le librerie Keras. Python and MySQL sono stati utilizzati rispettivamente, come linguaggio di programmazione e come sistema di gestione delle basi di dati. Su una GPU NVIDIA QUADRO P2000 ogni epoca richiede da 40 a 75 secondi per completare in funzione del tipo di architettura di RNC utilizzata. Mediamente ci vuole circa un'ora per elaborare 100 epoche per un singolo modello.

Nella Tabella 3 e nella Tabella 4 abbiamo riportato la somma dei dati della matrice di confusione per tutte le 12 partizioni degli studenti che hanno ottenuto l'indice F_1 migliore. Le somme della matrice di confusione relative ai dati della tabella *studente_A* in Tabella 3 e relative ai dati della tabella *studente_A* in Tabella 4.

In Tabella 5 e 6 si riportano i dati aggregati delle epoche che hanno ottenuto l'indice F_1 migliore su tutte e 12 le partizioni.



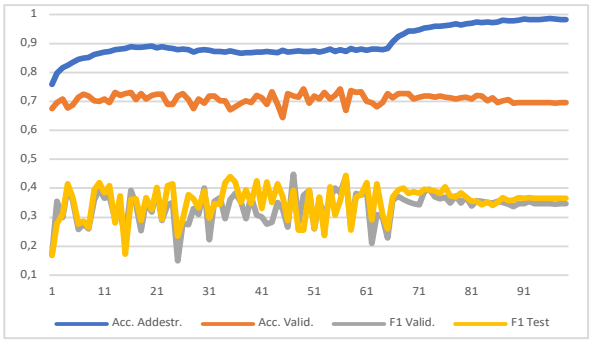


Figura 1. Nei grafi vengono riportati l'andamento dei seguenti valori: accuratezza dell'insieme di addestramento, l'accuratezza dell'insieme di validazione, l'indice F1 dell'insieme di validazione, e l'indice F1 dell'insieme di test per tutte le 100 epoche di addestramento dell'architettura RSNV2 e per la partizione P0 per la tabella studente_B. A sinistra: l'andamento relativo all'addestramento e all'anno di iscrizione 0. A destra: l'andamento relativo all'addestramento all'anno di iscrizione 1.

In Fig. 1 si riporta il grafico contenente l'andamento dei valori di accuratezza dell'insieme di addestramento, accuratezza dell'insieme di validazione, l'indice F_1 dell'insieme di validazione, e l'indice F_1 dell'insieme di test per tutte le 100 epoche di addestramento dell'architettura RSNV2 e per la partizione P_0. Entrambe le parti si riferiscono ai dati della tabella *studente_B*. La parte a sinistra è relativa all'anno 0 mentre quella a destra relativa all'anno 1. Si noti come l'indice F_1 sia notevolmente più basso nell'anno 1 rispetto all'anno 0.

In Fig. 2 si riporta il grafico contenente l'andamento dei valori di accuratezza dell'insieme di addestramento, accuratezza dell'insieme di validazione, l'indice F_1 dell'insieme di validazione, e l'indice F_1 dell'insieme di test per tutte le 100 epoche di addestramento dell'architettura RSNV2 e per la partizione P_0 dell'anno di iscrizione 1 della tabella *studente_A*.

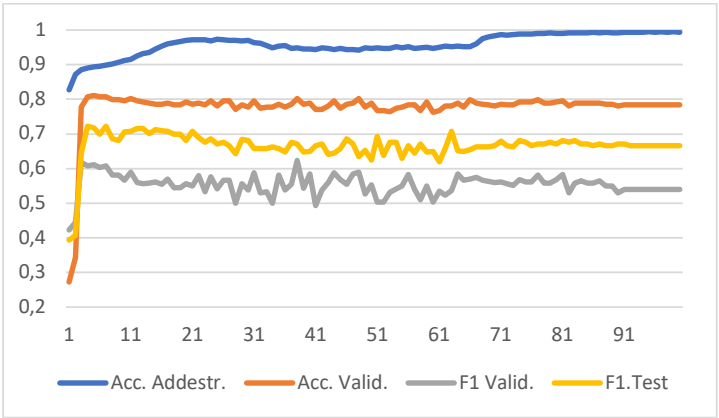


Figura 3. Nel grafo sono riportati l'andamento dell'accuratezza dell'insieme di addestramento, l'accuratezza dell'insieme di validazione, l'indice F1 dell'insieme di validazione, e l'indice F1 dell'insieme di test per tutte le 100 epoche di addestramento dell'architettura RSNV2 e per la partizione P0 e per all'anno di iscrizione 0 per la tabella studente_A.

Dall'andamento dei grafici e dalle matrici di confusione si nota che aggiungendo i dati delle carriere tutti gli indici relativi alla tabella *studente_A* (dall'anno di iscrizione 1 in poi) migliorano rispetto a quelli della tabella *studente_B*.

6. Conclusioni

Abbiamo realizzato diversi modelli predittivi dell'abbandono universitario utilizzando RNC. Questi modelli sono stati addestrati utilizzando i dati reali degli studenti del DSF di R3U iscritti tra il 2009 e il 2014. I risultati ottenuti sono molto incoraggianti considerando il fatto che molti dati significativi sono stati censurati, per motivi di privacy e considerando il fatto che importanti informazioni relative agli studenti non sono state raccolte (accuratamente) dagli uffici amministrativi di R3U. I dati ottenuti dai test chiaramente indicano che, mirate e maggiori informazioni possono migliorare la precisione e l'accuratezza dei modelli predittivi.

Riferimenti bibliografici

Allen, D. (1999). Desire to finish college: An empirical link between motivation and persistence. *Research in Higher Education*, 40, 461-85.

Anderman EM, Austin AC, Johnson DM. (2001). *The development of goal orientation*. See Wigfield & Eccles 2001.

ANVUR. (2018). *Rapporto sullo stato del Sistema universitario e della ricerca 2018*.

Astin, A. W. (1993). *What matters in college? Four critical years revisited*. San Francisco: Jossey Bass.

Bala, M. & 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, Barbaranelli C, Caprara GV, Pastorelli C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Dev.*, 72, 187-206.

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

Braxton, J. M., Sullivan, A. V., & Johnson, R. M. (1997). Appraising Tinto's theory of college student departure. In John C. Smart (Ed.), *Higher Education: Handbook of Theory and Research*, 12. New York: Agathon Press.

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade point average from the revised and recentered SAT I: Reasoning Test* (College Board Rep. No. 2000-1). New York: College Entrance Examination Board.

Burgalassi, M., Biasi, V., Capobianco, R., & 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.

Burgalassi, Marco, Biasi, V., Capobianco, R., & Moretti, G. (2017). The phenomenon of Early College Leavers. A case study on the graduate programs of the Department of Education of "Roma Tre" University. *Italian Journal Of Educational Research*, *0*(17), 105-126.

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

Carbone, V. & 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.

Cox, E. Orehovec, E. (2007). Faculty-Student Interaction Outside the Classroom: A Typology from a Residential College. *The Review of Higher Education*, 30, 4, Summer.

Delen, D. (2011). Predicting student attrition with *data mining* methods. *Journal of College Student Retention: Research, Theory & Practice*, 13(1), 17-35.

Des Jardins, S.L., Albourg, D.A., & Mccallan, P.B. (1999). An event history model of student departure. *Economics of education review*, 18, 375-90.

Di Pietro, G., & Cutillo, A. (2008). Degree flexibility and university drop–out: The Italian experience. *Economics of Education Review*, *27*(5), 546-555. https://doi.org/10.-1016/j.econedurev.2007.06.002

Fasanella A.,& Tanucci G. (2006). *Orientamento e carriera universitaria*. Milano: Franco Angeli.

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

Hu, N. B. (2002). *Measuring the Weight of High school GPA and SAT Scores with Second Term GPA to Determine Admission/ Financial Aid*. Paper Presented at the Annual Meeting of the Association for Institutional Research. 42nd, Toronto, Ontario, Canada.

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., & Rose, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(4), 333-353.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 1097-1105.
Kuncel, N.R., & Hezlett, S.A. (2007). Standardized tests predict graduate students' success. *Science*, 315, 1080-1081.
Kuncel, N. R., Credé, M., & Thomas, L. L. (2007). A comprehensive meta–analysis of the predictive validity of the Graduate Management Admission Test (GMAT) and undergraduate grade point average (UGPA). *Academy of Management Learning and Education*, 6, 51-68.
Kuncel, N.R., Hezlett, S.A., & Ones, D.S. (2001). A comprehensive meta–analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, 127, 162-181.
Kuncel, N. R., Hezlett, S. A., & 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, 148-161.
Larsen, M.R., Sommersel, H.B., & Larsen, M.S. (2013). *Evidence on dropout phenomena at universities*. Danish Clearinghouse for educational research Copenhagen.
Le, H., Casillas, A., Robbins, S., & Langley, R. (2005). Motivational and skills, social, and self– management predictors of college outcomes: Constructing the Student Readiness Inventory. *Educational and Psychological Measurement*, 65, 482-508.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436.
Lohfink, M.M., & Paulsen, M.B. (2005). Comparing the determinants of persistence for first–generation and continuing–generation students. *Journal of College Student Development*, 46, 409-428.
Malvestuto, F. M., Mezzini, M., & Moscarini, M. (2011). Computing simple-path convex hulls in hypergraphs. *Information Processing Letters*, 111(5), 231-234. https://doi.org/10.1016/j.ipl.2010.11.026
Manhães, L.M.B., da Cruz, S.M.S., & Zimbrão, G. (2014). The Impact of High Dropout Rates in a Large Public Brazilian University–A Quantitative Approach Using Educational Data Mining. *CSEDU* (3), 124-129.
Marshall, M. A., & Brown, J. D. (2004). Expectations and realizations: The role of expectancies in achievement settings. *Motivation and Emotion, 28*, 347-361.
Martinho, V. R. D. C., Nunes, C., & 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*, 159-166.
Mezzini, M. (2010). On the complexity of finding chordless paths in bipartite graphs and some interval operators in graphs and hypergraphs. *Theoretical Computer Science*, 411(7), 1212–1220. https://doi.org/10.1016/j.tcs.2009.12.017
Mezzini, M. (2011). Fast minimal triangulation algorithm using minimum degree criterion. *Theoretical Computer Science*, 412(29), 3775-3787. https://doi.org/10.1016/j.tcs.2011.04.022
Mezzini, M. (2012). Fully dynamic algorithm for chordal graphs with O(1) query-time and O(n2) update-time. *Theoretical Computer Science*, 445, 82-92. https://doi.org/10.1016/j.tcs.2012.05.002
Mezzini, M. (2016). On the geodetic iteration number of the contour of a graph. *Discrete Applied Mathematics*, 206, 211-214. https://doi.org/10.1016/j.dam.2016.02.012
Mezzini, M. (2018). Polynomial time algorithm for computing a minimum geodetic set in outerplanar graphs. *Theoretical Computer Science*, 745, 63–74. https://doi.org/10.1016/j.tcs.2018.05.032
Mezzini, M., & Moscarini, M. (2015). On the geodeticity of the contour of a graph. Discrete Applied Mathematics, 181, 209-220. https://doi.org/10.1016/j.dam.2014.08.028

Mezzini, M., & Moscarini, M. (2016). The contour of a bridged graph is geodetic. *Discrete Applied Mathematics*, 204, 213-215. https://doi.org/10.1016/j.dam.2015.10.007
Mezzini, M., Bonavolontà, G., & Agrusti, F. (2019). Predicting university dropout by using convolutional neural networks. *INTED2019*.
Milem, J.F., & Berger, J.B. (1997). A modified model of student persistence: Exploring the relationship between Astin's theory of involvement and Tinto's theory of student departure. *Journal of College Student Development*, 38, 387-400.
Moè, A., & De Beni, R. (2000). Strategie di autoregolazione e successo scolastico: Uno studio con ragazzi di scuola superiore e universitari. *Psicologia dell'Educazione e della Formazione*, *2*(1), 31-44.
Moretti, G., Burgalassi, M., & Giuliani, A. (2017, marzo). *Enhance students' engagement to counter dropping–out: a research at Roma Tre University*, 305-313. https://-doi.org/10.21125/inted.2017.0200
Mustafa, M. N., Chowdhury, L., & Kamal, M. S. (2012). Students dropout prediction for intelligent system from tertiary level in developing country. *2012 International Conference on Informatics, Electronics & Vision* (ICIEV), 113-118.
Nagy, M. & Molontay, R. (2018). Predicting Dropout in Higher Education Based on Secondary School Performance. *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*, 389-394.
Oppedisano, V. (2011). The (adverse) effects of expanding higher education: Evidence from Italy. *Economics of Education Review*, *30*(5), 997-1008. https://doi.org/10.1016/-j.econedurev.2011.04.010
Pascarella, E.T. (1980). Student–faculty informal contact and college outcomes. *Review of Educational Research*, 50 (4), 545-575.
Pascarella, E.T., & Terenzini, P.T. (1991). *How college affects students*. San Francisco: Jossey–Bass.
Pascarella, E.T., & Terenzini, P.T. (2005). *How College Affects Students: A Third Decade of Research* Jossey–Bass Higher & Adult Education.
Pereira, R.T., Romero, A.C., & Toledo, J.J. (2013). Extraction Student Dropout Patterns with *Data mining* Techniques in Undergraduate Programs. *KDIR/KMIS*, 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.
Pincus, F. (1980). The false promises of community colleges: Class conflict and vocational education. *Harvard Educational Review*, *50*(3), 332-361.
Pintrich, P. R. (2000). *The role of goal orientation in self–regulated learning*. See Boekaerts et al., 452-502.
Ramírez, P. E. & Grandón, E. E. (2018). Predicción de la Deserción Académica en una Universidad Pública Chilena a través de la Clasificación basada en Árboles de Decisión con Parámetros Optimizados. *Formación universitaria*, 11(3), 3-10.
Robins et al. (2004). Do Psychosocial and Study Skill Factors Predict College Outcomes? A Meta-Analysis. *Psychological Bulletin Copyright 2004 by the American Psychological Association*, 130, 2, 261-288.
Rosenbaum, J. E. (2004, Spring). *It's time to tell the kids: If you don't do well in high school, you won't do well in college (or on the job).* American Educator.
Rovira, S., Puertas, E., & Igual, L. (2017). Data–driven system to predict academic grades and dropout. *PLoS one*, 12(2), e0171207.
Siri, A. (2014). *Predicting students' academic dropout using artificial neural networks*. *Nova*.
Siri, A. (2015). Predicting students' dropout at university using artificial neural networks. *Italian Journal of Sociology of Education*, 7(2).
Sivakumar, S., Venkataraman, S., & 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.
Søgaard Larsen, M., & Dansk Clearinghouse for Uddannelsesforskning. (2013). *Dropout*

phenomena at universities: what is dropout? Why does dropout occur? What can be done by the universities to prevent or reduce it?: a systematic review. Danish Clearinghouse for Educational Research.
Solis, M., Moreira, T., Gonzalez, R., Fernandez, T., & Hernandez, M. (2018). Perspectives to Predict Dropout in University Students with Machine Learning. *2018 IEEE International Work Conference on Bioinspired Intelligence* (IWOBI), 1-6.
Stampen, J. O., & Cabrera, A. F. (1988). The targeting and packaging of student aid and its effect on attrition. *Economics of Education Review*, *7*(1), 29-46.
Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, *45*(1), 89-125.
Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition.* ERIC.
Tinto, V. (2010). From theory to action: Exploring the institutional conditions for student retention. *Higher education: Handbook of theory and research*,. Springer, 51-89.
Torrini, R. (2014). Rapporto sullo stato del sistema universitario e della ricerca 2013. *ANVUR–Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca, Roma*.
Troelsen, R., & Laursen, P.F. (2014). Is Drop–Out from University Dependent on National Culture and Policy? The Case of Denmark. *European Journal of Education*, *49*(4), 484-496.
Vossensteyn, J. J., Kottmann, A., Jongbloed, B. W. A., Kaiser, F., Cremonini, L., Stensaker, B., … Wollscheid, S. (2015). *Dropout and completion in higher education in Europe: main report*. https://doi.org/10.2766/826962
Yorke M. (2002). Formative Assessment – The Key To Richer Learning Experience In Year One. *Exchange*, 1, 12-13.