

Analisi preliminare per la definizione di un sistema predittivo per il completamento dei Mooc di EduOpen

Annamaria DE SANTIS, Katia SANNICANDRO, Claudia BELLINI, Tommaso MINERVA
Università degli Studi di Modena e Reggio Emilia, Reggio Emilia (RE)

Abstract

Obiettivo dello studio è costruire un sistema predittivo (e adattivo) che, in un ambiente digitale di apprendimento, stima (predice) il “completamento del corso” in funzione delle caratteristiche anagrafiche, motivazionali, comportamentali degli studenti.

La piattaforma individuata per la ricerca è EduOpen, portale italiano per l'erogazione di Mooc. In particolare, l'analisi si riferisce a 7 Mooc che differiscono fra loro per categoria, livello, lingua, durata, modalità di erogazione.

Nei corsi è stato somministrato un questionario nominale composto da 15 domande a risposta chiusa che ha consentito di identificare genere, età anagrafica, titolo di studio, profilo motivazionale e altre variabili di riferimento da studiare insieme a quelle inerenti i comportamenti degli utenti (svolgimento delle attività, numero di click) e correlare con le variabili dipendenti individuate nel completamento del corso e nella percentuale di completamento delle attività. Tale processo che permette di selezionare le variabili indipendenti dal set disponibile è il passo preliminare all'individuazione della forma funzionale (o parametrica) che descrive la relazione tra le modalità d'uso dell'ambiente digitale di apprendimento, il profilo degli utenti e le caratteristiche di erogazione dei corsi.

Keywords: MOOC, Learning Analytics, Course Completion, Users' Profile, Predictive Model

Introduzione

Lo sviluppo di modelli predittivi e l'identificazione di correlazioni fra azioni degli studenti e successo nella partecipazione ad attività formative, temi fra i più interessanti nel settore dei Learning Analytics (Gasevic *et al.*, 2019), sono utili per almeno tre motivi descritti da Josh Gardner e Christopher Brooks (2018): identificare celermente studenti “a rischio” che hanno bisogno di sostegno e definire azioni di supporto e intervento personalizzate ma allo stesso modo meno onerose da un punto di vista di impegno e di tempo per i professionisti coinvolti nella gestione dei MOOC; favorire l'apprendimento degli studenti attraverso l'implementazione di contenuti adattivi e suggerimenti di navigazione in tempo reale; comprendere i meccanismi legati al comportamento degli studenti nella frequenza ai corsi e nella navigazione di piattaforme didattiche.

Diverse sono le fonti da cui estrarre le informazioni, i significati del “successo” nella partecipazione di uno studente a un corso, le modalità di analisi dei dati. I click, i log, il tempo speso, così come le risposte a specifiche indagini di natura anagrafica o psicologica, la partecipazione ai forum, i risultati nelle prove di valutazione e gli artefatti prodotti sono alcuni dei dati più frequentemente disponibili e analizzati dai ricercatori; il successo è di volta in volta descritto come il completamento di tutte le attività, la certificazione, il superamento di tutte le prove di valutazione, l'engagement; la regressione lineare e logistica, l'albero delle decisioni, i network bayesiani e neurali rientrano fra i metodi usati per costruire modelli predittivi (Brown, 2012; Gardner & Brooks, 2018; Brooks & Thompson, 2017).

In questo contributo descriveremo le fasi iniziali di uno studio che intende definire un modello predittivo e adattivo per spiegare il completamento dei corsi a partire da variabili di natura demografica e comportamentale sulla piattaforma italiana dei MOOC, EduOpen, che, lanciata da circa 3 anni nel contesto italiano dell'alta formazione, raccoglie oggi 22 atenei.

Materiali e metodi

Lo studio condotto ha l'obiettivo di definire un sistema predittivo (e adattivo) che, in un ambiente digitale di apprendimento, stima (predice) il "completamento del corso" in funzione delle caratteristiche anagrafiche, motivazionali, comportamentali degli studenti. Il campione prende in esame 7 MOOC le cui caratteristiche sono descritte in Tab. 1. Le osservazioni raccolte riguardano 722 degli utenti/studenti iscritti ai corsi per i quali sono state analizzate congiuntamente sia le risposte fornite a un questionario nominale sia i report generati dalla piattaforma.

Il questionario, composto da 15 domande a risposta chiusa, indaga le caratteristiche anagrafiche e motivazionali degli studenti quali età, stato occupazionale e civile, formazione, impegno presunto nel completamento dei corsi, preconoscenze e motivazioni per concludere le attività e per abbandonare il MOOC.

Dai report della piattaforma, sono stati estratti i dati sui click per attività e sul completamento delle attività tracciate e dell'intero corso (il corso viene segnato come concluso in seguito al download dell'attestato di partecipazione).

Per ciascuno studente sono state quindi identificate 26 variabili che descrivono le caratteristiche personali dell'utente e le sue modalità di partecipazione al corso.

Le variabili indipendenti (24) sono state suddivise in quattro gruppi:

- 1) *profilo dell'utente* (GENDER, DEGREE, LANGUAGE, AGE, MARRIED, CHILDREN, TRAINING, WORKING, SECTOR, DIGITAL);
- 2) *engagement* (EFFORT, PRE.KNOWLEDGE, DROPOUT_TOT, DROPOUT_INT, DROPOUT_LEA, DROPOUT_NAV, MOTIVATION);
- 3) *comportamento dell'utente* (CLICKS_TRACKED, CLICKS_TOTAL);
- 4) *caratteristiche del corso* (CTUTORED, CCAT, CLANG, CHOUR, CLEVEL).

Le due considerate come variabili dipendenti nell'analisi sono la variabile binomiale del completamento dei corsi corrispondente al download dell'attestato (CERTIFICATE) e il rate di completamento delle attività tracciate presenti all'interno del MOOC da parte dello studente (CRATE).

Per la realizzazione dell'analisi preliminare, oggetto di questo contributo, sono stati utilizzati gli strumenti della statistica descrittiva monovariata e bivariata.

Corso	Categoria (CCAT)	Modalità di erogazione (CTUTORED)	Lingua (CLANG)	Durata in ore (CHOUR)	Livello (CLEVEL)
Collaboration and cooperation in industrial robotics	Scienze	Self-paced	EN	25	Intermedio
Cosa c'entra l'amore. Gli adolescenti e le relazioni di intimità in una ricerca emiliana	Scienze Sociali	Tutorato	IT	20	Beginner
Ethnobotany. The role of plants in our life	Scienze	Tutorato	EN	16	Intermedio
Introduzione al Sistema giuridico italiano	Scienze Sociali	Self-paced	IT	14	Beginner
Introduzione alla storia delle attività sportive	Scienze Umane	Self-paced	IT	14	Beginner
La violenza al Centro. Storie, sviluppi e pratiche di un Centro anti violenza italiano	Scienze Sociali	Tutorato	IT	20	Beginner
Physical synthesis of nanoparticles	Scienze	Tutorato	EN	20	Avanzato

Tabella 1 – Caratteristiche dei MOOC selezionati per lo studio.

Risultati

Descrizione del campione

Il campione può essere suddiviso in due gruppi di dimensioni quasi simili per genere: il 55.1% è di genere femminile, il 44.9% maschile. Se consideriamo la distribuzione in riferimento allo stato civile, risulta che il 42.4% è sposato/convivente e il 31.0% ha figli a carico. Studiando la distribuzione in base all'età, rileviamo che in prevalenza gli iscritti ai corsi sono adulti: la media della distribuzione è pari a 38 anni, la mediana si colloca nell'intervallo fra 31 e 35 anni e il 20.5% degli utenti ha meno di 25 anni.

La Figura 1 descrive la distribuzione degli studenti per livello di istruzione. Il 37.5% dei rispondenti all'indagine ha un titolo di istruzione secondaria superiore; tuttavia si ottiene una percentuale notevolmente superiore sommando il numero di quanti hanno un titolo uguale o superiore alla laurea di I livello (58.6%). Nella Figura 2 vediamo l'attuale partecipazione degli studenti a corsi di formazione: il 41.6% non partecipa ad attività formative strutturate e al contrario il 30.6% è attualmente impegnato in un corso di istruzione terziaria. Ben il 25.6% risponde a questa domanda con l'opzione "Altro". La Figura 3 presenta i risultati relativi allo stato occupazionale: il 53.6% ha un lavoro stabile, il 22.6% un lavoro precario e il 13.4% è in cerca di occupazione.

Gli studenti che scelgono le opzioni 4 e 5 per indicare il proprio livello di competenze digitali in una scala Likert a 5 livelli sono il 66.4%. Utilizzando la stessa scala, l'opzione 3 viene scelta dal 41.6% del campione per definire il proprio livello di preconoscenze sugli argomenti dei corsi, il 33.8% sceglie le opzioni 1 o 2 e il 24.6% le opzioni 4 e 5.

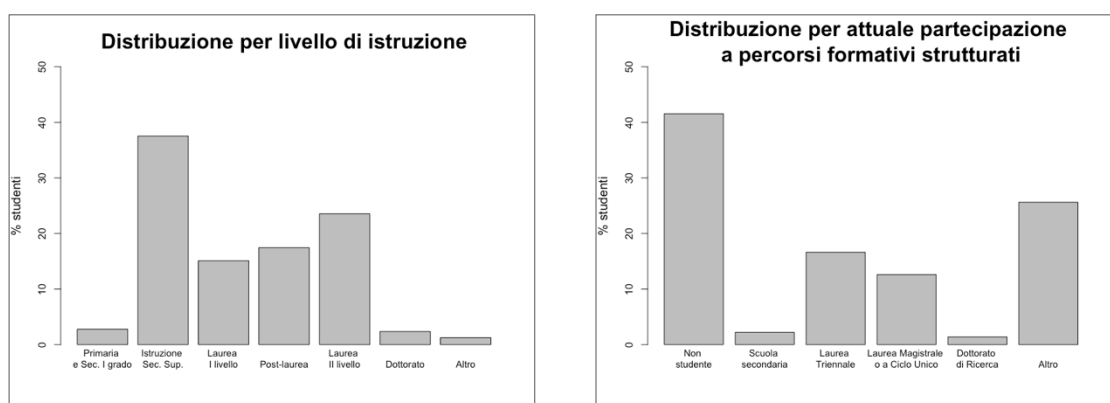


Figure 1 e 2 – Distribuzione del campione per livello di istruzione e attuale attività di formazione.



Figure 3 – Distribuzione del campione per stato occupazionale.

Nel gruppo di variabili relative all'*engagement* (si veda la Tab. 2), si verifica che gli studenti indicano mediamente 29 ore come quelle che prevedono di utilizzare per svolgere le attività dei corsi (EFFORT). I valori proposti dai manager del portale oscillano nei 7 MOOC fra le 14 e le 25 ore. I valori dei quartili

indicano che il 25% degli studenti ha dichiarato un impegno coerente con quello del range proposto dai manager dei corsi, il 25% sottostima l'impegno, il 50% lo sovrastima.

Oltre alle motivazioni per partecipare ai corsi (MOTIVATION) il cui valore in media si colloca a 23.0 in un range compreso fra 1.0 e 40.0 (SD=6.7), di questo gruppo fanno parte le quattro variabili relative alle motivazioni per abbandonare i corsi. La variabile DROPOUT_TOT raccoglie una misurazione complessiva di tali ragioni; le tre seguenti sono suddivise in base agli indicatori: interazione (DROPOUT_INT), attività didattiche (DROPOUT_LEA), navigazione (DROPOUT_NAV). Mentre la mediana per la seconda e la terza variabile è superiore al valore centrale della distribuzione, per la prima il valore è più basso del valore centrale e dunque meno rilevante per più del 50% degli studenti.

VARIABILE	MIN.	1° QU.	MEDIANA	MEDIA	DEV. ST.	3° QU.	MAX.
EFFORT	2.0	16.0	24.0	28.9	27.1	32.0	288.0
DROPOUT_TOT	3.0	21.0	26.0	25.2	7.6	30.0	45.0
DROPOUT_INT	0.0	2.0	4.0	4.1	2.1	6.0	10.0
DROPOUT_LEA	0.0	9.0	12.0	11.5	3.9	14.0	20.0
DROPOUT_NAV	0.0	4.0	6.0	6.0	2.6	8.0	10.0
MOTIVATION	1.0	18.3	23.0	23.04	6.7	27.0	40.0

Tabella 2 – Misure di tendenza centrale e quartili per il gruppo *engagement*.

Il numero medio di click per le attività di cui è tracciato il completamento è 3.5 (SD=5.5, range: 0.1-28.3); il valore scende a 2.6 (SD=2.3, range: 0.1-14.3) se consideriamo la totalità di attività e documenti presenti all'interno dei corsi (gruppo di variabili *comportamento degli utenti*).

Il 34.5% degli studenti ha completato i corsi selezionati scaricando l'attestato di partecipazione (CERTIFICATE) – il valore è superiore non solo rispetto a quello riportato nelle piattaforme internazionali di MOOC, ma anche rispetto alla media registrata sullo stesso EduOpen (pari a circa 25%). Si rileva però che il 44.3% degli utenti ha completato una percentuale di attività tracciate superiore al 90%. Considerando ancora il rate di completamento, il 28.8% degli utenti ha svolto meno del 10% delle attività tracciate nel corso. Risulta quindi che la variabile CRATE ha una distribuzione all'incirca bimodale (si veda la Figura 4) nella quale distinguiamo studenti che non hanno frequentato affatto i corsi e altri che avendo completato più della metà delle attività, concludono il percorso con l'acquisizione dell'attestato.

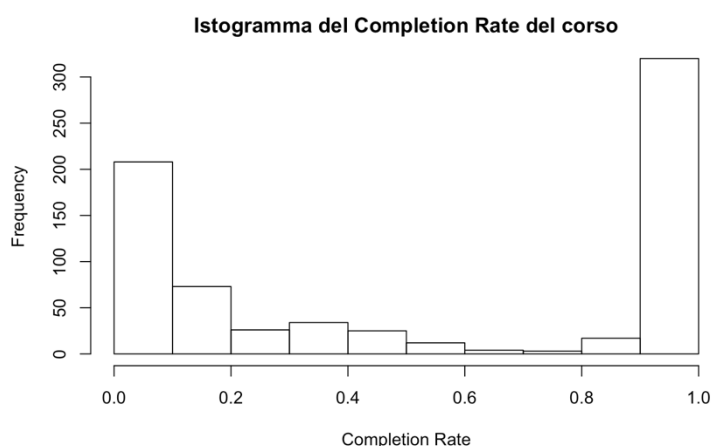


Figura 4 – Distribuzione del campione per percentuale di completamento del corso.

Correlazione lineare fra le variabili indipendenti

Nei gruppi di variabili relativi al *completamento*, al *comportamento degli utenti* e alle *caratteristiche dei corsi* le correlazioni, come ci si aspettava, sono molto alte per la stretta relazione che esiste fra le variabili misurate. Più irregolari sono i valori di ρ considerando le variabili relative al *profilo* e all'*engagement* degli studenti (Tab. 3). Risulta una lieve correlazione fra genere, età e stato civile,

impegno e caratteristiche dei corsi in base alla quale si può affermare che le donne sono con minor frequenza sposate e con figli, più giovani rispetto agli uomini del campione, mostrano un livello inferiore di prenoscenze e maggiore interesse per i corsi tutorati. All'aumentare dell'età, aumentano le possibilità che gli utenti abbiano conseguito un più alto titolo di studio e siano sposati/conviventi con figli, ugualmente aumenta l'interesse per corsi in autoapprendimento e più brevi.

Valori più alti del coefficiente di correlazione si registrano fra le variabili del gruppo *engagement*: non solo fra le variabili dedicate alle motivazioni per il dropout dove ρ è compreso fra 0.33 e 0.91 ma anche in riferimento alle variabili PRE.KNOWLEDGE e MOTIVATION. Maggiori conoscenze degli argomenti del corso comportano una maggiore motivazione nella partecipazione allo stesso e motivazioni più alte si traducono in maggiori aspettative di interazione con colleghi/docenti e di qualità di risorse e portali. Gli studenti che dichiarano di avere più elevate competenze digitali, affermano anche di avere più prenoscenze sugli argomenti dei corsi, motivazioni più elevate, maggiore interesse nel confronto con colleghi e docenti.

	CRATE	CERTIFICATE	GENDER	DEGREE	LANGUAGE	AGE	MARRIED	CHILDREN	TRAINING	WORKING	SECTOR	DIGITAL	EFFORT	PRE.KNOWLEDGE	DROPOUT_TOT	DROPOUT_INT	DROPOUT_LEA	DROPOUT_NAV	MOTIVATION	CLICK_TRACKED	CLICK_TOTAL	CLUTORED	CCAT	CLANG	CHOUR	CLEVEL
CRATE	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-CERTIFICATE	0.77	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GENDER	-0.13	-0.20	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DEGREE	-0.01	-0.03	0.08	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LANGUAGE	-0.09	-0.13	0.00	0.10	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AGE	0.12	0.16	-0.11	0.26	-0.10	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MARRIED	0.04	0.07	-0.13	0.13	-0.04	0.50	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CHILDREN	0.10	0.11	-0.15	0.08	-0.08	0.51	0.64	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TRAINING	0.04	0.04	0.08	0.13	0.02	0.01	-0.07	-0.07	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
WORKING	-0.01	-0.02	0.07	-0.11	-0.04	-0.33	-0.18	-0.21	0.08	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SECTOR	-0.02	0.00	0.04	-0.10	0.10	-0.34	-0.17	-0.16	0.06	0.21	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DIGITAL	0.01	0.03	-0.09	0.08	0.12	-0.03	0.05	0.04	0.01	0.02	-0.05	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-
EFFORT	0.03	0.00	0.10	-0.02	0.01	-0.02	-0.01	0.00	0.02	-0.04	0.02	0.04	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-
PRE.KNOWLEDGE	0.09	0.09	-0.16	0.06	-0.07	0.03	0.06	0.12	0.11	0.02	0.00	0.29	0.11	1.00	-	-	-	-	-	-	-	-	-	-	-	-
DROPOUT_TOT	-0.02	0.04	0.04	0.09	0.01	-0.01	-0.01	0.01	-0.02	-0.02	0.01	0.05	-0.03	0.10	1.00	-	-	-	-	-	-	-	-	-	-	-
DROPOUT_INT	0.04	0.07	-0.02	-0.02	0.06	-0.11	-0.06	0.01	0.02	0.05	0.06	0.12	0.04	0.18	0.68	1.00	-	-	-	-	-	-	-	-	-	-
DROPOUT_LEA	-0.02	0.02	0.01	0.10	0.00	0.00	0.00	0.02	-0.06	-0.05	0.01	0.06	-0.02	0.08	0.91	0.51	1.00	-	-	-	-	-	-	-	-	-
DROPOUT_NAV	0.02	0.05	0.06	0.09	0.02	0.00	-0.02	0.03	-0.01	0.02	-0.03	-0.03	0.03	0.76	0.33	0.54	1.00	-	-	-	-	-	-	-	-	-
MOTIVATION	0.17	0.15	-0.02	0.03	0.01	-0.06	-0.07	-0.04	0.17	-0.01	0.04	0.15	0.03	0.32	0.26	0.34	0.19	1.00	-	-	-	-	-	-	-	-
CLICK_TRACKED	0.77	0.64	0.01	-0.07	-0.11	0.08	0.01	0.06	0.01	0.02	0.03	-0.02	0.00	0.04	-0.01	0.05	-0.01	0.00	1.00	-	-	-	-	-	-	-
CLICK_TOTAL	0.86	0.73	-0.07	-0.05	-0.11	0.16	0.05	0.08	0.03	-0.03	0.01	-0.03	-0.01	0.04	-0.02	0.02	-0.03	0.03	0.14	0.92	1.00	-	-	-	-	-
CLUTORED	0.02	-0.08	0.21	0.03	-0.02	-0.20	-0.15	-0.15	0.01	0.08	0.10	0.03	0.03	-0.01	0.03	0.00	0.04	-0.01	0.09	0.17	0.04	1.00	-	-	-	-
CCAT	0.03	0.04	0.00	-0.10	-0.04	0.09	0.13	0.13	-0.02	0.03	-0.06	-0.01	-0.05	0.08	-0.01	0.05	-0.02	-0.03	-0.06	0.10	0.04	-0.57	1.00	-	-	-
CLANG	-0.03	-0.04	0.00	0.10	0.04	-0.09	-0.13	-0.13	0.02	-0.03	0.06	0.01	0.05	-0.08	0.01	-0.05	0.02	0.03	0.06	-0.10	-0.04	0.57	-1.00	1.00	-	-
CHOUR	0.03	-0.05	-0.04	0.02	-0.01	-0.11	-0.10	-0.02	-0.04	0.03	0.05	0.04	0.06	0.01	0.03	0.05	0.06	-0.04	0.07	0.09	-0.03	0.34	-0.49	0.49	1.00	-
CLEVEL	-0.03	-0.04	0.00	0.10	0.04	-0.09	-0.13	-0.13	0.02	-0.03	0.06	0.01	0.05	-0.08	0.01	-0.05	0.02	0.03	0.06	-0.10	-0.04	0.57	-1.00	1.00	0.49	1.00

Tabella 3 – Matrice di correlazione lineare delle variabili dello studio.

Correlazione lineare fra variabili dipendenti e indipendenti

La correlazione di ciascuna variabile con le due variabili risposta (CERTIFICATE e CRATE) assume nella maggior parte dei casi valori prossimi allo 0 e compresi fra -0.20 e 0.17, dato che comporta l'ipotesi che la relazione stabilita fra le variabili dipendenti e indipendenti non abbia una forma lineare. Si notano valori di ρ più elevati nel gruppo *profilo* per le variabili GENDER, LANGUAGE, AGE, CHILDREN e nel gruppo *engagement* per PRE.KNOWLEDGE e MOTIVATION. Questi risultati ci dicono che gli uomini, di lingua italiana, adulti e con figli hanno possibilità lievemente più elevate di completare i corsi o di svolgere percentuali più alte di attività in essi contenute. Inoltre, prenoscenze e motivazioni contribuiscono positivamente al raggiungimento di tale obiettivo.

Rispetto al quadro appena descritto, fanno eccezione le due variabili indipendenti che descrivono il *comportamento degli studenti* all'interno dei corsi in funzione dei click (CLICK_TRACKED e CLICK_TOTAL): in questo caso i valori assunti da ρ sono compresi fra 0.64 e 0.86. Essendo in una correlazione lineare più stretta con le variabili dipendenti, solo gli unici che sembrano effettivamente influenzarne l'andamento.

Conclusioni

Lo studio, di cui in questo contributo vengono descritte solo le fasi preliminari, ha lo scopo di individuare le relazioni esistenti fra le caratteristiche e i comportamenti degli studenti di 7 MOOC erogati sulla piattaforma EduOpen e il completamento dei corsi al fine di definire un sistema predittivo e adattivo.

Le analisi finora condotte ci hanno restituito innanzitutto una descrizione del campione che tiene conto di elementi anagrafici, occupazionali ma anche motivazionali e comportamentali. Gli studenti coinvolti in questa indagine risultano essere adulti, con un titolo di istruzione terziaria e una situazione lavorativa stabile, più del 60% dei rispondenti afferma di avere ottime competenze digitali e la stessa percentuale definisce buone/ottime le conoscenze pregresse sugli argomenti dei corsi. Dallo studio abbiamo acquisito alcune informazioni rispetto al fenomeno del completamento dei corsi che è di interesse nelle ricerche sull'open e l'online education e, a maggior ragione, su EduOpen, dove i completamenti risultano evidentemente più alti rispetto alle medie internazionali. Nel campione osservato, la percentuale di completamento dei corsi supera infatti il 34%. Tuttavia il rate di completamento delle attività tracciate assume una distribuzione bimodale, dove oltre a una percentuale di poco inferiore al 30% di utenti che non svolgono affatto attività nel corso dopo essersi iscritti, se ne registra una superiore al 40% di iscritti che hanno svolto più del 90% di attività nel corso e che, superata la metà delle attività tracciate nel corso, proseguono fino ad arrivare al completamento.

I bassi valori del coefficiente di correlazione ci permettono di identificare lievi legami fra le coppie di variabili fornendo informazioni rispetto alla relazione (o mancata relazione) fra caratteristiche dei corsi e degli utenti e completamento dei corsi. I legami fra le variabili non sembrano essere di tipo lineare e probabilmente sarà necessario nel proseguimento dello studio utilizzare algoritmi genetici per determinare le trasformazioni matematiche più appropriate (Paterlini & Minerva, 2010). Il coefficiente ρ assume valori più elevati solo fra le variabili dipendenti (CRATE, CERTIFICATE) e il numero di click (CLICK_TRACKED, CLICK_TOTAL) che potremmo assimilare alle azioni che lo studente mette in atto all'interno del portale, all'uso che fa delle risorse e alle attività che svolge all'interno del corso. Tale elemento porta l'attenzione sulle modalità di progettazione e realizzazione dei MOOC, sulla qualità e sulla validità di ciascuna risorsa.

La ricerca proseguirà con un'analisi *stepwise* che, utilizzando l'Asymptotic Information Criterion (Akaike, 1969; 1978) per la scelta delle variabili, permetterà di selezionare un modello di regressione lineare per rappresentare la relazione fra le variabili coinvolte nello studio.

Riferimenti bibliografici

Akaike A. (1969), Statistical predictor identification, *Annals of the Institute of Statistical Mathematics*, 22, 203-217.

Akaike A. (1978), Bayesian analysis of the minimum AIC procedure, *Annals of the institute of Statistical Mathematics*, 30, part A, 9-14.

Brooks C. & Thompson C. (2017), *Predictive Modelling in Teaching and Learning*, in: Lang C., Siemens G., Wise A. & Gašević D. (eds.), *Handbook of Learning Analytics*. 61-68, SOLAR: Society for Learning Analytics Research.

Brown, M. (2012). *Learning Analytics: Moving from Concept to Practice*. (EDUCAUSE Learning Initiative Briefs). Louisville, CO: EDUCAUSE.

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Model User-Adap Inter*, 28(2), 127-203.

Gašević D., Tsai Y.S., Dawson S. & Pardo A. (2019), How do we start? An approach to learning analytics adoption in higher education, *The International Journal of Information and Learning Technology*.

Paterlini, S., & Minerva, T. (2010). Regression Model Selection using Genetic Algorithms. In V., Munteanu, R., Raducanu, G., Dutica, A., Croitoru, V.E., Balas, & A. Graviut (Eds.), *Recent Advances in Neural Networks, Fuzzy Systems & Evolutionary Computing* (pp. 19-28). USA: WSEAS Press.