



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE DELLA RICERCA

Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

Come sfruttare gli Educational Data ? Un inquadramento di usi e metodologie di analisi

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Gioia Taraborrelli, Matteo Farne (2022). Come sfruttare gli Educational Data ? Un inquadramento di usi e metodologie di analisi. *INDUZIONI*, Anno 2021(62/63), 27-40 [10.19272/202100902002].

Availability:

This version is available at: <https://hdl.handle.net/11585/916031> since: 2023-02-18

Published:

DOI: <http://doi.org/10.19272/202100902002>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

Come sfruttare gli Educational Data? Un inquadramento di usi e metodologie di analisi

Gioia Taraborrelli, Matteo Farnè

Dipartimento di Scienze Statistiche, Università di Bologna

This article provides an overview of the most relevant statistical methods used to explore educational data and to derive key analytical insights from them. We describe two broad categories of statistical methods: those pertaining to Educational Data Mining and Learning Analytics. The former category gathers the methods aimed at uncovering hidden patterns from educational data, with the aim to improve and facilitate the task of all the actors involved in the education process, like students, instructors, and administrators. The latter category gathers the methods specifically oriented to measure how much student learn and how. A detailed explanation of the most popular statistical methods into both categories is offered to the reader, as well as an analysis of their similarities and differences. We conclude that Learning Analytics is typically more concerned with pedagogical aspects involving human judgment, while Educational Data Mining is more oriented to identify in automated way optimal practices to improve the outcomes of the learning process from student feedback.

Keywords: Educational Data Mining, Learning Analytics, longitudinal data, high dimension, data privacy

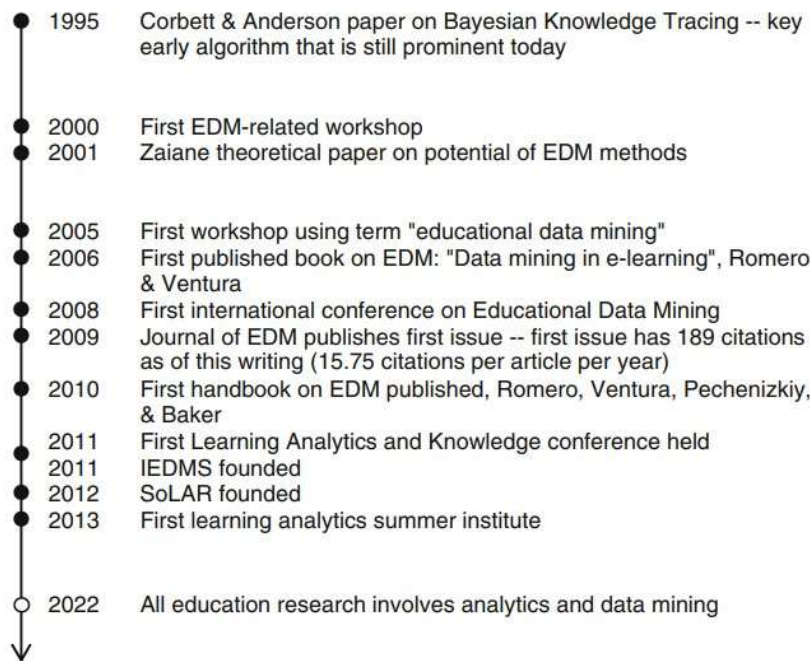


Figura 2 - Linea temporale dell'Educational Data Mining (Baker e Inventado, 2014).

Secondo Baker e Yacef (2009), le aree di ricerca dell'EDM sono molteplici, tra cui l'apprendimento individuale sui software scolastici, l'apprendimento collaborativo supportato dai computer, i test che si adattano in base al livello di abilità dello studente (*computer adaptive test*) e i fattori che causano la bocciatura o l'insuccesso degli studenti.

Gli stessi autori hanno individuato le seguenti aree di applicazione:

- miglioramento degli *student model*, che contengono le informazioni sulle caratteristiche di uno studente (il suo livello di conoscenza, le sue motivazioni e le sue attitudini). Grazie agli *student model*, che riescono a modellare le differenze individuali, si riesce a rispondere alle esigenze individuali di apprendimento e a migliorare di conseguenza il percorso di apprendimento di uno studente. Essi sono utili perché riescono a prevedere il livello di conoscenza e la performance futura di uno studente e permettono ai ricercatori di studiare i fattori che influenzano le scelte in un contesto di apprendimento (Baker, 2010);
- ricerca e miglioramento dei *domain model*, che sono dei modelli che riguardano un dominio di studio specifico (nel nostro caso si tratta del mondo della didattica) e rappresentano la struttura di conoscenza in tale ambito. Essi vengono spesso combinati con modelli psicometrici e vengono utilizzati a scopi predittivi (ad esempio, cercano di prevedere se le azioni di un soggetto saranno corrette o errate);
- studio del supporto alla didattica fornito dai software per l'apprendimento;

- ricerca di evidenza empirica per capire in maniera più approfondita i fattori chiave che influenzano l'apprendimento.

Fino ad ora sembra che gli attori principali dell'EDM siano unicamente gli studenti, anche se effettivamente vengono coinvolte moltissime figure in tutto il processo educativo. Gli studenti sono coloro che beneficeranno *in primis* di tutte le innovazioni nell'ambito educativo, ma bisogna ricordare che “dietro le quinte” ci sono anche ricercatori e insegnanti, nonché le istituzioni scolastiche. Romero e Ventura (2010) hanno identificato coloro che sono i principali stakeholder in questa disciplina e i modi in cui le tecniche di Data Mining possono aiutare a migliorare l'intero sistema scolastico.

Stakeholder	Miglioramenti possibili
Studenti Chi apprende (in generale)	<ul style="list-style-type: none"> • Personalizzazione della didattica • Suggestire attività che possono migliorare l'apprendimento • Suggestire esperienze di apprendimento interessanti • Supporto personalizzato in base allo studente • Consigliare corsi da seguire o libri da leggere
Insegnanti Tutor Educatori	<ul style="list-style-type: none"> • Analizzare il comportamento e l'apprendimento degli studenti • Individuare gli studenti che hanno bisogno di supporto • Prevedere la performance degli studenti • Classificare gli studenti in gruppi • Scoprire pattern regolari (o irregolari) nei dati • Trovare gli errori più frequenti commessi dagli studenti • Determinare le attività più efficaci • Migliorare l'adattamento e la personalizzazione dei corsi

<p>Ricercatori Coloro che sviluppano i corsi di studio</p>	<ul style="list-style-type: none"> • Valutare il materiale didattico • Migliorare l'apprendimento degli studenti • Valutare i contenuti del corso e la loro efficacia • Costruire <i>student model</i> e <i>tutor model</i> • Confrontare le tecniche di data mining per suggerire la tecnica più utile per svolgere una determinata azione • Sviluppare metodi di data mining finalizzati alla didattica
<p>Scuole Università Organizzazioni Scuole private</p>	<ul style="list-style-type: none"> • Valorizzare il processo decisionale nelle università • Semplificare il processo decisionale • Suggerire corsi specifici per ogni tipo di studente • Trovare soluzioni parsimoniose per migliorare i voti e mantenere il numero di studenti inalterato • Ammettere gli studenti che potranno avere buoni risultati
<p>Amministratori School District Administrators Network Administrators System Administrators</p>	<ul style="list-style-type: none"> • Sviluppare il miglior modo per organizzare le risorse (umane e materiali) al servizio della didattica • Utilizzare le risorse attuali in maniera più efficiente • Migliorare i programmi scolastici • Valutare gli insegnanti e i corsi di studio • Fissare dei parametri per migliorare il sito web scolastico

Tabella 1.1 – Stakeholder del sistema educativo e modi per migliorarlo grazie a metodi di Data Mining. Rielaborazione da Romero e Ventura (2010).

La personalizzazione della didattica e i sistemi in grado di adattarsi alle esigenze di chi apprende hanno una loro utilità; tuttavia, questo potrebbe creare delle problematiche legate al trattamento dei dati, alla privacy e all'etica. Ad esempio, su una qualsiasi piattaforma e-Learning, può essere raccolta un'enorme varietà di dati, dai semplici dati di navigazione di un utente fino alle risposte date in un test presente sulla piattaforma. Quindi, coloro che analizzano tali dati dovrebbero dare suggerimenti

ad hoc solo ai diretti interessati e dovrebbero agire in modo tale da non danneggiare nessuno. I legislatori dovrebbero valutare la validità dei modelli predittivi che riguardano gli studenti e dovrebbero fornire evidenza per la previsione e per le azioni decise dagli algoritmi. Altri problemi legati alla privacy potrebbero sorgere quando un'istituzione scolastica non possiede un database di sua proprietà e fa affidamento su terze parti per la raccolta, l'immagazzinamento, l'analisi dei dati (che riguardano informazioni sui propri studenti e insegnanti). In tal caso bisognerebbe creare un identificativo che garantisca la loro privacy, dato che questi dati potrebbero essere utilizzati in molti tipi di analisi; in dataset che contengono un elevato numero di dati longitudinali, queste analisi potrebbero portare alla divulgazione involontaria di informazioni sensibili (Bienkowski et al., 2012).

Esistono anche problemi giuridici legati all'EDM in quanto il GDPR (*General Data Protection Regulation*) limita i processi decisionali automatizzati: il paragrafo 1 dell'articolo 22 afferma che un soggetto "ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona"¹, tranne nel caso in cui tale processo sia necessario per la conclusione di un contratto, oppure esista una legge dell'Unione Europea o di quel determinato Stato che autorizza il processo oppure il soggetto esprima il suo consenso esplicitamente sul trattamento dei suoi dati. In queste tre situazioni, colui che dovrà trattare i dati avrà il dovere di illustrare come funziona tale processo e quali sono i suoi criteri e le sue conseguenze, deve dimostrare come questi metodi siano necessari al raggiungimento di un determinato obiettivo, deve rendere note le basi giuridiche e deve fare in modo che il soggetto interessato possa esprimere la propria opinione o abbia la possibilità di contestare nel caso in cui non fosse d'accordo (Redazione Utopia, 2020).

Volendo guardare l'Educational Data Mining dal punto di vista epistemologico², Aristizabal (2016) ha fornito un'analisi critica riguardo i suoi metodi di osservazione e sperimentazione. In particolare, egli ha criticato l'approccio positivista dell'EDM, che consiste nella semplice deduzione di conoscenza a partire da informazioni empiriche, e ha consigliato di includere l'interpretazione del ricercatore nelle analisi dei dati, perché fino ad ora molte ricerche si sono basate su dati di navigazione, o di altro tipo, raccolti su piattaforme e-Learning e non dalle interazioni dirette o "faccia a faccia" all'interno di una qualsiasi classe. Perciò, secondo l'autore, l'Educational Data Mining

¹ Fonte: Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati, L. 119/34 pubblicato nella GUUE del 4.5.2016.

² L'epistemologia, in senso più ampio, può essere vista come la filosofia della scienza, che studia dal punto di vista critico i modi per raggiungere la conoscenza scientifica.

dovrebbe andare oltre i processi automatizzati, che potrebbero generare risultati validi solo per un particolare gruppo di studenti e quindi non sarebbero così generalizzabili, e dovrebbe inventare nuovi metodi per analizzare i dati provenienti da classi “reali”. Per fare ciò, i ricercatori dovrebbero trovare il modo di creare software e tecniche al servizio degli insegnanti delle scuole in modo che questi ultimi possano mettere in pratica nella vita di tutti i giorni ciò che la teoria dell’EDM suggerisce.

3. I metodi statistici più utilizzati nell’Educational Data Mining

Baker e Inventado (2014) hanno classificato i vari metodi statistici applicati nell’Educational Data Mining in quattro categorie, basandosi su una serie di articoli scientifici.

1. **Modelli predittivi.** Viene sviluppato un modello che può spiegare una variabile dipendente tramite una qualche combinazione delle variabili indipendenti. I modelli predittivi si dividono in modelli di classificazione e di regressione. I primi comprendono alberi decisionali, *random forests*, regole decisionali, regressione per passi e regressione logistica; i secondi comprendono la nota regressione lineare e gli alberi di regressione. In questo ambito, sono poche le applicazioni delle reti neurali artificiali e delle macchine a supporto vettoriale a causa dell’elevata variabilità nei dati e dei numerosi regressori. In generale, vengono applicati per individuare quali tipi di comportamento determinano la scelta di una certa scuola superiore oppure per prevedere il livello di conoscenza di uno studente. È molto frequente la stima del livello di conoscenza latente tramite reti Bayesiane per strutture di conoscenza complesse, Bayesian Knowledge Tracing e approcci basati sulla regressione logistica.
2. **Structure Discovery.** Questi algoritmi sono finalizzati a trovare la struttura nei dati senza avere conoscenze pregresse sui dati (apprendimento non supervisionato). Tra i vari metodi possiamo trovare il clustering, l’analisi fattoriale e la *domain structure discovery*.
3. **Relationship Mining.** Questi metodi si propongono di scoprire le relazioni tra le variabili in un dataset che contiene un elevatissimo numero di variabili. Tra i metodi più comuni troviamo le regole di associazione, il *sequential pattern mining*, il *correlation mining* e il *casual data mining*.
4. **Discovery with Models.** Anche se non è un ambito molto esplorato nel Data Mining classico, nel contesto dell’EDM viene usato in questo modo: vi è un modello iniziale che descrive un fenomeno e viene sviluppato attraverso la previsione o il clustering; la variabile dipendente di tale modello diventa poi un regressore di un nuovo modello predittivo.

Gli autori affermano che durante i primi anni di diffusione dell’EDM il Relationship Mining rappresentava l’argomento più diffuso negli articoli scientifici, anche se col tempo è sempre più

diminuita la sua trattazione; la previsione e il clustering rimangono fin dalle origini tra i metodi più importanti; negli ultimi anni sono aumentate le ricerche nel contesto della stima della conoscenza latente e della *domain structure discovery*.

Villanueva et al. (2018) hanno condotto uno studio sulle varie tecniche di DM applicate all'EDM, facendo riferimento ad un corpus composto da più di 100 articoli scientifici. Di seguito possiamo osservare una tabella tratta dal loro articolo, nella quale vengono elencati i vari metodi di DM e le rispettive applicazioni.

Technique \ Domain	Dropping out or Retention Analysis	VLO or VLE Analysis	Performance and students evaluation Analysis	Generation of Educational Recommendations	Learning pattern Identification	Students patterns Identification	Students related Prediction
Correlation Analysis		1				1	
Decision Trees	5	3	8	2	2	6	2
Regression Trees			1				
Markov Chains				1			
Classification	4	2	4		3	1	3
Clustering		7	3	5	3	9	2
Differential Sequence Mining						1	
Sequential Patterns		4	1	1	7	3	2
Bayesian Networks		2	1		1	1	6
Neural Networks	1	2	2		1		5
Association rules		8	1	7	14	9	1
Linear regression					1		1

Figura 3 - Tecniche di DM applicate all'Educational Data Mining (Villanueva et al., 2018).

Tra i metodi di DM troviamo: l'analisi della correlazione, gli alberi decisionali, gli alberi di regressione, le catene di Markov, algoritmi di classificazione, il *Clustering*, il *Differential Sequence Mining method*, il *Sequential Patterns Mining method*, le reti Bayesiane, le reti neurali, le regole di associazione e la regressione lineare. Tra le applicazioni, invece, vi sono: l'analisi della probabilità di abbandonare la scuola, l'analisi del *Virtual Learning Object* (VLO) o del *Virtual Learning Environment* (VLE), l'analisi della performance e della valutazione degli studenti, generare consigli sui contenuti scolastici, identificazione dei pattern di apprendimento, identificazione dei "tipi" di studenti, fare previsioni legate agli studenti.

Si può notare che gli alberi decisionali sono tra gli algoritmi più usati (in particolar modo per analizzare la performance degli studenti); sono seguiti dalle regole di associazione (particolarmente utilizzate nell'identificazione di pattern di apprendimento e dei tipi di studenti), dal clustering (identificazione dei tipi di studenti), dai *Sequential Patterns methods*, dalle reti neurali e Bayesiane. I metodi meno usati sono l'analisi della correlazione, la regressione lineare, le catene di Markov, gli alberi di regressione e i *Differential Sequence Mining methods*.

Le applicazioni più comuni sono l'identificazione di pattern (di apprendimento e di studenti), l'analisi del VLO e del VLE, l'analisi della performance degli studenti e le previsioni riguardanti gli studenti. Quelle meno comuni riguardano la generazione di consigli per gli studenti e la previsione degli studenti che hanno probabilità maggiori di abbandonare la scuola.

4. Learning Analytics

Il Learning Analytics (LA) è definito come il procedimento di analisi dei dati di apprendimento (dall'inglese *educational data*) che include la misura, la raccolta, l'analisi e il resoconto dei dati sugli studenti e sul contesto scolastico, per capire e ottimizzare l'apprendimento e l'ambiente in cui gli studenti imparano (Lang et al., 2017). Il Learning Analytics si occupa dell'analisi di dati preesistenti e le sue tecniche sono in grado di gestire i Big Data: grazie alla diffusione dei Learning Machine Systems sono stati accumulati sempre più dati sugli studenti e sull'apprendimento (Ferguson, 2014). Reimann (2016) afferma che i metodi utilizzati dal Learning Analytics si adattano ad un elevato numero di dati, longitudinali, provenienti da contesti diversi e che riguardano diversi livelli di apprendimento (nel tempo).

Il Learning Analytics è nato “ufficialmente” all'inizio del 2011, quando un gruppo di ricercatori ha tenuto una conferenza per la prima volta una conferenza sull'argomento in Canada, chiamata “First Conference on Learning Analytics & Knowledge” (LAK 2011), con l'obiettivo di definire e circoscrivere questo tema di ricerca emergente che si focalizzava sulla comprensione dell'apprendimento da parte degli studenti tramite l'utilizzo del Machine Learning, del Data Mining e dei metodi di visualizzazione. Grazie a questa conferenza venne poi fondata nel 2012 la Society of Learning Analytics Research (SoLAR) (Joksimović et al., 2019).



Figura 7 - Logo della SoLAR (<https://www.solaresearch.org/about/>)

Avella et al. (2016) hanno condotto uno studio sui metodi più utilizzati nel Learning Analytics, sui loro benefici e sulle sfide da loro poste nel contesto universitario. Gli autori hanno analizzato 34 articoli scientifici e hanno dedotto che nel Learning Analytics (LA) sono molto comuni le tecniche di visualizzazione dei dati e la social network analysis. In più, dall'Educational Data Mining il LA “prende in prestito” modelli predittivi, metodi di clustering, *Relationship Mining* e *Discovery with Models*. I benefici che ne derivano sono l'offerta di corsi mirati, lo sviluppo di *curricula* migliori, migliori risultati nell'apprendimento degli studenti, l'apprendimento personalizzato, il

miglioramento della performance degli insegnanti, più opportunità alla fine del percorso universitario e la valorizzazione della ricerca scientifica nell'ambito didattico. I problemi che si dovrebbero risolvere riguardano il tracciamento, la raccolta, la valutazione e l'analisi dei dati, la mancanza di connessioni con le scienze dell'apprendimento, l'ottimizzazione degli ambienti di apprendimento e gli ambiti dell'etica e della privacy.

5. Educational Data Mining e Learning Analytics: punti in comune e differenze

È opportuno sottolineare che, nonostante l'ambito di applicazione comune nell'ambito della didattica, EDM e Learning Analytics sono due discipline separate che spesso tendono a sovrapporsi. Secondo Siemens e Baker (2012), le differenze tra EDM e LA sono da riscontrarsi nella loro metodologia; l'EDM è più orientato sugli aspetti tecnici e tecnologici; infatti, esso ricerca metodi per scoprire le strutture nascoste nei dati scolastici e per riconoscere e proporre soluzioni personalizzate automaticamente in base ai bisogni dello studente. Il Learning Analytics è invece focalizzato sugli aspetti educativi: infatti i ricercatori si propongono di supportare nel miglior modo possibile gli stakeholder (intesi come studenti, insegnanti, ricercatori e amministratori) nell'analisi dei dati scolastici e di informare gli insegnanti sulle eventuali difficoltà degli studenti circa un argomento specifico, al fine di supportarli.

Educational Data Mining	Learning Analytics
Metodi automatici per analizzare i dati	Centralità del giudizio umano
Approccio riduzionistico	Approccio olistico
Adattamenti automatici	Intervento umano
Apprendimento come argomento di ricerca	Aspetti educativi come argomento di ricerca

Tabella 1.3 – Sintesi delle differenze tra EDM e LA. Rielaborazione tratta da Baker e Inventado (2014).

Sia nell'Educational Data Mining che nel Learning Analytics vengono utilizzati gli stessi metodi, anche se nell'EDM vengono preferiti metodi automatici e nel LA si dà più importanza all'interpretazione umana e alla visualizzazione dei dati. I metodi predittivi vengono utilizzati in entrambi gli ambiti: nel LA, in particolare, ci si focalizza maggiormente sulla classificazione e sulla regressione. Sono molto utilizzati anche i metodi di Discovery Structure, come il clustering. I metodi di *Relationship Mining* e di *Discovery with Models* sono più comuni nell'EDM. Nel Learning Analytics sono più frequenti invece metodi di text mining e text analysis (Baker, Inventado, 2014).

Lemay et al. (2021) hanno svolto una *topic modeling analysis* di articoli scientifici, pubblicati tra il 2015 e il 2019, che includevano ricerche empiriche su EDM e LA al fine di sottolineare i loro punti

in comune e le loro differenze. Essi hanno raggruppato le applicazioni comuni di EDM e LA in quattro dimensioni:

- analisi dell'apprendimento supportata dai computer, che include la previsione del tasso di abbandono della scuola da parte degli studenti, la previsione e la valutazione della performance degli studenti;
- analisi predittiva supportata dai computer, che include l'apprendimento di gruppo ed individuale;
- analisi del comportamento degli studenti supportata dal computer, che si prefigge di modellare gli schemi di apprendimento degli studenti;
- analisi della visualizzazione dei dati supportata dai computer, come metodi basati su grafici o mappe.

Dal *topic modeling* è emerso che i paper riguardanti il Learning Analytics fossero focalizzati sull'insegnamento, sulla performance degli studenti e sull'apprendimento in un corso: in particolare, è stato notato che l'inclusione della parola "corso" servisse a sottolineare maggiormente l'applicazione delle teorie del LA alle tecniche di insegnamento e di apprendimento. I topic più frequenti nei paper riguardanti l'EDM sono "performance" e "sistema": questo ci fa capire che vi sono meno applicazioni nell'ambito delle tecniche di insegnamento rispetto al LA. Inoltre, gli autori hanno sottolineato brevemente l'importanza della privacy e dell'etica in quanto, nei paper da loro analizzati, essi non erano dei topic molto comuni: secondo loro bisognerebbe approfondire i temi che riguardano la raccolta trasparente dei dati, la gestione dei protocolli di sicurezza nel momento in cui i dati vengono immagazzinati in un database, la proprietà dei dati e il consenso degli studenti/utenti.

Bibliografia

- Aristizabal, A. (2016). An Epistemological Approach for Research in Educational Data Mining. *International Journal of Education and Research*, Vol. 4 No. 2 February 2016.
- Avella, J. T.; Kebritchi, M.; Nunn, S. G.; Kanai, T. (2016). *Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review*. *Online Learning*, Volume 20 Issue 2, June 2016.
- Baker, R. (2010). *Data Mining for Education*. In McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education* (3rd edition), vol. 7, pp. 112-118. Oxford, UK: Elsevier.
- Baker, R. S.; Inventado, P. S. (2014). *Educational Data Mining and Learning Analytics*. *Learning Analytics* (pp.61-75).
- Baker, R. S. D.; Yacef, K. (2009). *The state of educational data mining in 2009: A review and future visions*. *Journal of Educational Data Mining*, vol. 1, no. 1, 2009.
- Bienkowski, M.; Feng, M.; Means, B. (2012). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*. U.S. Department of Education, Office of Educational Technology.
- Ferguson, R. (2014). *Learning analytics: fattori trainanti, sviluppi e sfide*. *TD Tecnologie Didattiche*, 22(3), 138-147.
- Joksimović, S.; Kovanović, V.; Dawson, S. (2019). *The Journey of Learning Analytics*. *HERDSA Review of Higher Education*, Vol. 6, 2019.
- Lang, C., Siemens, G., Wise, A., & Gasevic, D. (2017). *Handbook of learning analytics*. SOLAR, Society for Learning Analytics and Research. New York, NY: SOLAR In Romero, C.; Ventura, S. (2020). *Educational data mining and learning analytics: An updated survey*. *WIREs Data Mining Knowl Discov*. 2020;10:e1355.
- Lemay, D. J.; Baek, C.; Doleck, T. (2021). *Comparison of learning analytics and educational data mining: A topic modeling approach*. *Computers and Education: Artificial Intelligence*, Volume 2, 2021.
- Redazione Utopia (25/11/2020). L'algoritmo che valuta gli studenti: Educational Data Mining. *Utopia the Software*. <https://www.utopiathesoftware.com/blog-post/algoritmo-valuta-studentieducational-data-mining-gdpr> [Data di accesso: 07/05/2022]

Reimann, P. (2016). *Connecting learning analytics with learning research: the role of design-based research*. *Learning: Research and Practice*, 2(2), 130–142.

Romero, C.; Ventura, S. (2010). *Educational Data Mining: A Review of the State-of-the-Art*. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*. 40(6), 601-618.

Villanueva, A.; Moreno, L. G.; Salinas, M. J. (2018). *Data mining techniques applied in educational environments: Literature review*. *Digital Education Review - Number 33*, June 2018.