

IL RUOLO DEI BIG DATA NELLE STRATEGIE DI APPRENDIMENTO

Massimiliano Giacalone, Sergio Scippacercola¹

Dipartimento di Scienze Economiche e Statistiche-Università degli Studi di Napoli
"Federico II"-Complesso Universitario di Monte S. Angelo-Via Cinthia, 26-80126-Napoli
massimiliano.giacalone@unina.it

¹Dipartimento di Economia, Management, Istituzioni - Università degli Studi di Napoli
"Federico II"-Complesso Universitario di Monte S. Angelo-Via Cinthia, 26-80126-Napoli
sergio.scippacercola@unina.it

Il lavoro richiama i concetti fondamentali sui Big data, evidenziando, in particolare, le applicazioni delle tecniche di analisi e delle tecnologie a supporto nell'ambito dell'e-learning. Gli enormi volumi di dati disponibili al giorno d'oggi stanno diventando protagonisti dell'e-learning, ad esempio per identificare in modo proattivo le esigenze di apprendimento degli allievi. Lo studio confronta i Big Data con il Data Mining, le tecniche di e-learning e di learning-analytics. Sono accennati i principi di etica e di privacy da osservare e sono, infine, riportati i principali vantaggi derivanti dall'adozione dei Big Data nel settore educativo.

1. Introduzione

Quando si fa riferimento ai Big Data si pensa a insiemi di dati metrici, oppure indifferentemente non metrici, caratterizzati da elevata numerosità, e da una certa dinamicità in termini di cambiamento in valore, nel loro dominio di definizione. Pertanto, insiemi di immagini, e-mail, dati GPS, o informazioni desunte da siti web (come accessi, permanenze, etc.) possono essere definiti *Big Data* [Snijders, et. al., 2012].

Una delle caratteristiche fondamentali dei Big Data è l'eterogeneità delle fonti dei dati: si tratta di insiemi di dati, o frequentemente di flussi dinamici di 'metadati', provenienti da database eterogenei [Rezzani, 2013]. Ad esempio, non soltanto censimenti, rilevazioni, interviste, o questionari; ma anche raccolte informative da Internet, da reti telefoniche, da satelliti, o da reti di trasporto (es. i dati di treni, aerei, autostrade, varchi autostradali e stradali controllati, etc.), possono far parte di uno stesso insieme di dati.

Una definizione comune di Big Data è quella offerta da Doug Laney [Laney, 2001], che si basa **sul paradigma delle tre VVV** (VOLUME, VELOCITÀ, VARIETÀ):

- **Volume:** si stima che entro il 2020 verranno generati 35 mila miliardi di gigabyte di dati; per quanto attiene alla enumerazioni delle basi di Big Data, si è proceduto progressivamente ad estendere l'unità di misura via via che procedevano in estensione i volumi medi di grandezza in essere, pervenendo ad oggi a ordini di grandezza volumetrici espressi in 'Zettabyte', pari a un miliardo di Terabyte e Yottabyte pari a un trilardo (10^{21}) di byte.

- **Velocità:** una volta estratti, i dati devono essere analizzati tempestivamente, affinché non divengano obsoleti, e quindi inutili per prendere una "decisione". La VELOCITÀ di acquisizione e accesso ai dati richiesti è dunque fondamentale. Basta pensare che non è infrequente la necessità di acquisire dati 'in tempo reale' (ad esempio, accessi ai siti, ai motori di ricerca in ambito Internet, o dati di share in ambito televisivo), per elaborarli con cadenza giornaliera, se non intra-giornaliera.

- **Varietà:** i dati hanno natura fortemente eterogenea (es., testi, immagini, video, ricerche sul web, operazioni finanziarie, email, post su blog e social network, etc.), ed ogni formato necessita di un trattamento dedicato. Tale caratteristica dei Big Data può richiedere operazioni di *scaling* o classificazioni convenzionali (ad esempio catalogare delle immagini per data cronologica, oppure per scala cromatica o secondo un'altra scala ordinativa) [Manyika et al., 2011].

Alcuni studiosi suggeriscono di aggiungere alla definizione dei Big data altre due V:

- **Variabilità:** i dati vanno contestualizzati, in quanto il loro significato può variare a seconda del contesto.

- **Viralità:** la crescita dei Big data avviene in maniera esponenziale, a macchia d'olio.

Tali peculiari caratteristiche e specifiche richiedono che, rispetto alla memorizzazione, i database costituenti i Big Data, siano sia di tipo strutturato che non strutturato, e siano espressi su differenti scale di misura o siano anche di tipo qualitativo.

Fortunatamente, i progressi compiuti nel frattempo dalla ricerca scientifica e tecnologica in ambito hardware e software hanno garantito soddisfacenti prestazioni in termini di efficienza, di velocità di accesso ai Big Data e di potenza ed efficacia elaborativa.

Sia la costruzione dei dataset e dei pattern di Big Data, che l'elaborazione numerica ivi applicabile, richiede usualmente algoritmi di 'calcolo parallelo'. In questo contesto si rivela utile l'adozione di 'algoritmi genetici' capaci di operare una riunione di dati anche non metrici provenienti da fonti dinamiche coagenti nel processo di formazione dei dataset. Detti algoritmi possono utilizzare metodologie di selezione categoriale o anche di tipo probabilistico (selezione di 'Boltzmann') [Koza, 1992] [Wright, 1991].

I Big Data sono nati a causa della enorme e massiccia proliferazione di dati elementari provenienti da più fonti e diffusi su tutto il territorio. Spesso i Big Data sono confusi con le semplici tracce digitali di attività umane mediate dalla tecnologia ICT, come sono le memorizzazioni di accesso a servizi che vengono detti i "log" di servizio. Basti pensare alle numerose telefonate o messaggi

scambiati continuamente. I “log” sono infatti delle registrazioni che consistono in un identificativo del richiedente e talvolta di un breve contenuto testuale o multimediale e talora, come nel caso delle telefonate da unità mobili, anche del contenuto spaziale. Anche i log, comunque, possono fare parte di un sistema di Big Data.

Usare i Big Data per misurare e comprendere fenomeni sociali invece è un approccio multidisciplinare il cui scopo è sviluppare metodi appropriati finalizzati a conoscenza, previsione e supporto alle decisioni [Snijders et al., 2012] in realtà di crescente complessità caratterizzate appunto dalla disponibilità di masse di dati di diversa natura (numeri, simboli, testi, immagini, dati relazionali ecc.) provenienti dalle fonti più disparate (*Social Network*) ad esempio: le tracce delle targhe rilevate dai sistemi tutor, le tracce dei biglietti degli autobus (origine–destinazione), il traffico dei messaggi tra studenti, il traffico delle telefonate web, il traffico delle registrazioni telepass, etc.

Si possono usare le interrogazioni di Internet (*Google Trends*), come informazioni ausiliarie, termini come "lavoro", "offerte di lavoro" e simili da Google Trends al fine di produrre stime anticipate e di previsione (*nowcasting e forecasting*). Si riporta, a titolo di esempio, una rete di comunità estratta mediante tracciamenti (*GPS*) di rete con telefoni mobili in Belgio per individuare i sistemi locali di lavoro. Si tratta di una Rete con circa 3 milioni di utenti e l'analisi si è sviluppata su 118 milioni di nodi cioè su un bilione di link (1000 miliardi) [Blondel et. al., 2008] (Fig. 1).

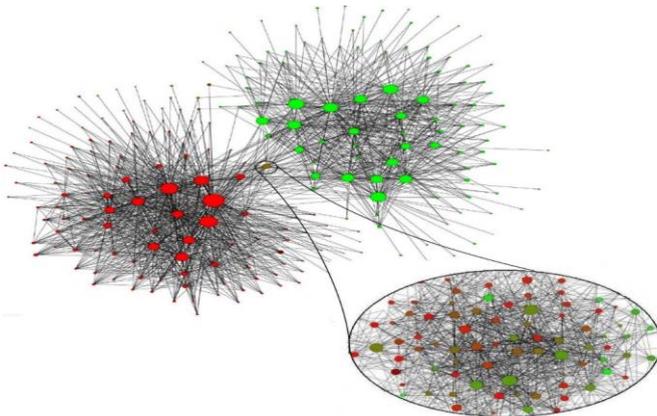


Fig. 1 - Rete di comunità estratta mediante tracciamenti di rete con telefoni mobili in Belgio [Blondel et. al., 2008].

Obiettivo del presente lavoro è di disegnare lo stato dell'arte relativo al ruolo che oggi stanno assumendo i Big Data nell'e-learning. Nel secondo paragrafo si confrontano le metodologie del Data Mining, già utilizzato da tempo, e le nuove frontiere aperte dai Big Data. Nel terzo paragrafo sono evidenziati i rapporti tra i Big Data e le tecniche di e-learning. Learning-analytics con i Big Data sono presentati nel quarto paragrafo. Nel quinto sono discussi i principi di etica e di

privacy nell'era dei Big Data con qualche cenno sui costi attuali. Nel sesto paragrafo sono descritti i principali vantaggi dell'uso dei Big Data nel campo educativo. Nell'ultimo paragrafo sono tratte alcune conclusioni. Infine, in Appendice sono riportati alcuni software per l'e-learning e per i Big Data.

2. Data Mining e Big Data

Il termine Data Mining fa parte della Business Intelligence, ed indica il procedimento di esplorazione ed analisi di un insieme di dati per individuare eventuali regolarità, estrarre nuova conoscenza e ricavare regole ricorrenti significative. Obiettivo principale del "Data Mining" è di "estrarre informazioni" utili da un database e trasformarle in una struttura di dati (pattern) per ulteriori utilizzi di indagine. Tra le principali applicazioni del Data Mining possiamo evidenziare la descrizione riassuntiva dei dati, le associazioni e le correlazioni, le classificazioni e le analisi di evoluzione (regolarità dei dati che cambiano nel tempo). Le tecniche del Data Mining sono adottate in diversi campi: nella statistica, nelle scienze della formazione, in economia, in medicina, etc.

Evidenti sono le analogie riscontrabili tra i "Big Data" e il "Data Mining". Questo ultimo potrebbe essere considerato il "**vecchio Big Data**" perché risponde almeno in parte a due delle caratteristiche dei Big Data che sono il Volume e la Velocità, manca però la terza V (la varietà) in quanto nel Data Mining spesso si estrae conoscenza solo mediante i Database o i Datawarehouse e Data Mart che sono di tipo statico retrospettivo a differenza dei Big Data che sono costantemente aggiornati con una frequenza elevata e diventano sempre più accurati e precisi con il passare del tempo.

Un altro aspetto interessante dei Big Data che lo differenzia dal Data Mining è la **diversità strutturale**. Alcuni dati hanno un formato ben definito, nel modo classico di file/record/campi, come, ad esempio, nelle transazioni registrate in un Database; altri dati possono essere di tipologia molto differente, come una raccolta di testi su un blog, o tabelle, o immagini, o una registrazione audio, o un video. Dal punto di vista delle architetture e della ingegnerizzazione dei dataset e delle strutture dei dati, i modelli più recenti dei Big Data sono basati su metodologie ad elevata scalabilità, e su soluzioni di tipo **NoSQL** [Vaish, 2013]. Si intende per NoSQL un insieme di tecnologie formanti un nuovo sistema di gestione dei dati diverso dal tradizionale **RDBMS**, in quanto non viene utilizzato il modello relazionale, non ha uno schema esplicito e il sistema è progettato per funzionare subito e bene in cluster.

Inoltre, quelle che rappresentano le problematiche precipue dell'analisi e del trattamento dei dati della business intelligence - dati errati, dati mancanti, dati duplicati, integrazione delle basi di dati - risultano spesso secondarie quando si opera con i Big Data, dove spesso esigenze quali la pulizia dei dati attendibili da quelli inattendibili o scarsamente veritieri risultano invece prioritarie. Peraltro, tipicamente i Big Data presentano condizioni di ridondanza informativa, nel senso che per conseguire la informazione 'sufficiente' ed 'esauriente' contenuta

nei dati è preferibile operare con approccio campionario e conseguentemente inferenziale previamente alla sedimentazione aggregativa e ai task che preludono alla creazione effettiva delle basi dati e dai dataset a partire dalle basi 'grezze' di Big Data incipienti [Manoochehri, 2013].

Sinteticamente, l'aggregazione delle fonti numeriche promiscue viene affrontata operando su flussi dati in parallelo (approccio 'map') poi sottoposti a trattamenti di riduzione, filtraggio e 'pulizia' dei dati eliminando quelli non veritieri o inutili (dati 'garbage') prima di operare le aggregazioni e le riorganizzazioni finali in dataset [Reiss et al., 2012].

3. Big Data e e-Learning

L'impatto che hanno i Big Data in ambito didattico - sia con riferimento all'insegnamento, che all'apprendimento - è rilevante, non soltanto nella progettazione dei moduli didattici, ma anche in tema di affinamento di obiettivi di apprendimento già predefiniti [Gutierrez-Santoz et al., 2012]. I Big Data possono essere utilizzati in molteplici settori e l'e-learning è uno di questi. Ogni qual volta i discenti interagiscono con il contenuto di un corso, infatti, essi producono dati. Oggi siamo in grado di tracciare e raccogliere questi dati non solo attraverso i learning management systems (LMS), ma anche tramite i *social networks* e ogni altro media.

Accanto alla usuale 'valutazione di fine-corso', per mezzo di questionari di gradimento proposti ai discenti, cresce ed assume rilevanza l'esigenza di acquisire *in tempo reale* informazioni sempre più dettagliate ed organizzate sui vari ambiti di valutazione della didattica. Ad esempio, gli accessi ('le visite') alle pagine Web costituiscono basi informative che possono essere acquisite *on line* assieme ad altri dati, per comporre pattern utili alla valutazione della didattica.

Dai Big Data, i docenti responsabili dell'e-Learning possono ricevere informazioni per rendere più efficace l'insegnamento, o per correggere eventuali difetti. Ad esempio, gli accessi ai siti web, i dati prelevati dai social network, il contenuto delle ricerche sul web, ed i moduli di apprendimento *on line*, possono costituire Big Data utili per valutare la fruizione informativa da parte dei discenti, ed i loro comportamenti nella fase di apprendimento.

Una prerogativa interessante è data dalla possibilità, utilizzando appositi programmi software o *tool* di potere immediatamente scartare i dati non utili dal punto di vista informativo. L'uso di modelli matematici e di metodi statistici sui dati dell'e-Learning, una volta organizzati gli stessi in basi di dati o 'metadati', consente di produrre modelli di comprensione o anche di previsione utili all'affinamento o alla semplice valutazione dei metodi didattici [Chatti et al., 2012].

Un altro approccio di utilizzo di Big Data, è valutare differenti parametri della formazione didattica prefissando per ciascuna variabile opportuni 'valori-soglia' ovvero 'livelli-target' da raggiungere per ritenere conseguiti gli obiettivi-formativi. [Siemens et al., 2011].

In un articolo del 2014 dal titolo “Big Data in eLearning: The Future of eLearning Industry”, Christopher Pappas [Pappas, 2014] ha elencato a tal proposito 5 benefici che si possono trarre dall’analisi dei dati relativi alla fruizione di un corso e-learning:

1. L’analisi dei dati consente di individuare quale tipologia di attività didattica è più efficace per il raggiungimento degli obiettivi del corso.

2. Diventa possibile individuare i miglioramenti del percorso didattico. Ad esempio, se un grande numero di discenti impiega un tempo eccessivo per completare un certo modulo, significa che quel modulo deve essere reso più snello e fruibile.

3. E’ possibile monitorare quali sono i moduli più visualizzati o i link più condivisi con gli altri discenti.

4. I dati relativi risultanti dalle tracce del discente sono immediatamente disponibili e non occorre attendere i risultati dei test di valutazione finale per conoscere la situazione. In tal modo i docenti possono ottenere un quadro complessivo dei comportamenti dei discenti e possono ottimizzare la strategia di apprendimento quasi in tempo reale.

5. Sulla base dei dati è possibile fare previsioni sui successi e sui fallimenti dei discenti e sviluppare i corsi in maniera tale che gli studenti abbiano sempre la possibilità di ottenere il miglior risultato possibile [Pappas, 2014].

In sintesi, il vantaggio principale della raccolta e dell’analisi dei Big Data nell’e-learning, sta soprattutto nella possibilità di ricavare informazioni utili per personalizzare l’esperienza formativa sulla base delle esigenze e degli stili di apprendimento dei discenti.

4. Big Data e Learning analytics

Il termine *learning analytics* identifica un settore di rilievo all’interno del *Technology-Enhanced Learning* emerso negli ultimi anni ed è strettamente collegato a diverse discipline come la *Business Intelligence*, la *Web analytics* e l’*Educational Data Mining (EDM)*. Con il termine *learning analytics* ci si riferisce alla misurazione, alla raccolta, all’analisi e alla presentazione dei dati sugli studenti e sui loro contesti ai fini della comprensione e dell’ottimizzazione dell’apprendimento e degli ambienti in cui ha luogo [Baker et al., 2014][Ferguson, 2012, 2014].

In Ferguson è riportato, a titolo di esempio, il progetto Signals [Ferguson, 2014], elaborato dalla Purdue University che esplora grandi dataset e applica test statistici al fine di prevedere, durante lo svolgimento dei corsi, quali studenti rischiano di rimanere indietro. L’obiettivo è quello di produrre *actionable intelligence*, guidando gli studenti verso risorse appropriate e spiegando loro come usarle. Un semaforo mostra agli studenti se le cose stanno andando bene (verde), oppure se sono stati classificati come ad alto rischio (rosso) o a rischio moderato (giallo) [Arnold, 2010].

I risultati riportati appaiono promettenti anche se il sistema considerato dei dati e del software non possono essere del tutto assimilabili ad un sistema di

Big Data. Da un punto di vista tecnologico learning analytics è una nascente disciplina e le sue connessioni con i Big Data, nonostante alcune significative proposte in College americani [Picciano, 2012], resta ancora da sviluppare.

5. Etica, privacy e costi

Le attività di raccolta dati dell'e-Learning hanno differenti ricadute nel delicato ambito delle problematiche afferenti il trattamento dei dati personali, definiti 'sensibili' ai sensi della normativa e del complessivo sistema di vincoli e responsabilità, regolate dalle leggi sulla privacy.

Rilevanti sono gli aspetti etici derivanti dalla gestione dei contenuti di dati personali degli studenti, con riferimento anche alle possibilità di diffusione, condivisione e fruizione delle informazioni 'sensibili' riguardanti gli stessi discenti. Ad oggi esistono delle criticità che hanno già indotto le organizzazioni professionali a una seria riflessione sui contenuti e sulla delimitazione dei confini operativi delle attività di e-Learning, allo scopo di pervenire alla realizzazione di piani di raccolta e interventi 'legali' capaci cioè di qualificare i vari aspetti delle attività in essere (la raccolta, la conservazione, la gestione, la fruizione, la pubblicazione dei dati personali) in osservanza di 'best practices' predefinite [Slade et al., 2013].

Uno dei temi più importanti è la comunicazione degli intenti e delle finalità di una raccolta di dati per l'e-Learning, così da conseguire preliminarmente l'autorizzazione e legittimazione nel rispetto della normativa vigente, rendendo chiara la comunicazione ai discenti interessati a ciascuno dei quali richiedere esplicito consenso al trattamento dei propri dati personali. Se vi sono 'stakeholder' o committenti esterni della raccolta dei dati, allo stesso modo le organizzazioni e i professionisti che realizzino studi o analisi di dati in ambito e-learning devono svolgere tutti gli atti di comunicazione e produrre relativa contrattualistica conforme alle normative vigenti in ordine al trattamento dei dati personali, sia per quanto attiene alla eventuale tenuta che per quanto riguarda le modalità di diffusione dei dati personali desunti o evinti dalle attività in parola. Un altro aspetto non secondario è quello della conservazione in sicurezza e dell'accessibilità dei dati personali in server dotati di procedure, protocolli e standard di sicurezza attiva e passiva, così come previsto dalle normative già da anni in vigore e dagli standard internazionali di sicurezza (ISO,EN) [Corposanto et al., 2014].

Per quanto attiene alla conservazione e all'accessibilità dei dati, le tecnologie a supporto dei Big Data sono altamente affidabili e scalabili a basso costo. Ad esempio, Hadoop [Hadoop, 2014] è un sistema adeguato ai Big Data, perché consente di immagazzinare enormi volumi di dati per poi processarli quando più è opportuno. Hadoop consente, inoltre, la distribuzione dei dati su più nodi, riducendo i costi computazionali e di storage per la memorizzazione e l'analisi dei Big Data, e mascherando i fallimenti hardware. È stato stimato da Zedlewski [Zedlewski et al., 2003] che il costo di un sistema di gestione dei dati basato su Hadoop, considerando i costi hardware, software, e altre spese,

ammonta a circa \$ 1.000 per terabyte, ovvero da un quinto a un ventesimo del costo di altre tecnologie di gestione dei dati.

6. I vantaggi dei Big Data nel campo educativo

I Big Data sono attualmente utilizzati da varie società per corsi di formazione ed anche in ambito universitario: con l'aiuto dei Big Data, è possibile guardare i discenti ed esaminare le tracce dei loro percorsi individuali. Ad esempio, è possibile individuare le pagine web su cui i discenti si intrattengono di più o su cui trovano maggiore difficoltà di apprendimento, quelle che rivisitano spesso, e determinare i giorni e gli orari in cui lavorano di più, etc. Pertanto, i Big Data ci aiutano a comprendere i veri modelli comportamentali dei discenti, molto di più di quanto oggi avviene mediante l'istruzione tradizionale.

Questi modelli portano ad informazioni interessanti su cosa e come imparano, contribuendo in tal modo a prendere decisioni più informate circa i programmi di apprendimento e ad individuare difetti di progettazione dei corsi. Tuttavia, la vera forza dei Big Data sta nel loro potere per aiutare previsioni o prevedere scenari per adottare misure preventive. Ad esempio, con l'aiuto di Big Data, è anche possibile effettuare previsioni su quali siano i concetti che risultano ostici agli studenti, gli argomenti che generano confusione e difficoltà nell'apprendimento.

Big Data oggi si presenta come una piattaforma efficace che stravolge il modo tradizionale con cui l'e-learning è nato. Mediante i Big Data è possibile progettare piani di apprendimento più personalizzati e idonei per gli studenti. Il monitoraggio diventa l'elemento principe usato dagli educatori. Vengono utilizzati gli stessi risultati conseguiti dagli studenti per migliorare la loro formazione.

Con gli strumenti di oggi a disposizione dei discenti (mobile, tablet, smart phone, tecnologie cloud-based, etc.) l'infrastruttura è ormai consolidata. I *Data Analytics* permettono di ottenere un quadro molto più preciso del monitoraggio rispetto al passato con i metodi tradizionali finora usati. Se il docente prevede un eventuale abbandono, ha tutto il tempo di attuare azioni di recupero personalizzato onde evitare che lo studente lasci il corso.

7. Conclusioni

I Big Data sono utilizzabili per l'e-learning, e la diffusione degli ambienti virtuali per l'apprendimento (*Virtual Learning Environment*) avranno un impatto sullo sviluppo e la diffusione di questi sistemi, che potranno offrire nuove tecniche di apprendimento *on line*.

Al giorno d'oggi assistiamo alla confluenza dei Big Data, il Data Mining, la Statistica, la Matematica, le Scienze del computer, i Data Warehouse, l'Intelligenza artificiale, e le reti neurali in un nuovo paradigma, che prende il nome di **Data Science**, e che promette di rivoluzionare il mondo, investendo

tutti i settori, dall'assistenza sanitaria, sino al mondo accademico. In tale prospettiva, Data Science rivoluzionerà anche l'insegnamento.

Il paradigma nato con la Data Science consiste nell'estrarre dati di ogni tipologia esistenti "nel mondo", applicare idonei modelli, ottenere analisi descrittive dei fenomeni, reimmettere risultati nel circuito mondiale e così via, perfezionando sempre di più la *conoscenza* utile al domino applicativo. Da Data Science, nasce una nuova figura professionale, il **Data Scientist**, che ha il compito di analizzare i dati per fornire al committente le informazioni utili ad assumere decisioni.

Il Data Scientist lavorerà anche per l'e-learning, disegnando dall'analisi dei Big Data nuove e più efficaci strategie di insegnamento: egli dovrà sapere elaborare dati ed essere competente in analisi statistica, per stabilire quali trasformazioni effettuare sul percorso formativo e per suggerire le scelte da compiere, in base agli obiettivi dell'apprendimento.

Pertanto, il ruolo dei Big Data è non solo nella capacità di poter gestire *velocemente grandi volumi differenti di dati di vario tipo*, ma è dato dall'opportunità che queste tecnologie offrono per nuove e notevoli applicazioni, anche in ambito educativo.

Appendice - Software per l'e-Learning e i Big Data

Sono state messe a punto varie applicazioni software di tipo interattivo per l'e-Learning a scopo predittivo o di valutazione delle performance didattiche. Tra queste si citano: Apache HBASE, BEESTAR Insight, DOCEBO, LITMOSLMS, HALOGEN Software.

Per la gestione di sistemi Big Data si citano: HADOOP, MAPREDUCE, HDFS, YARN, CASSANDRA, MONGODB.

Riferimenti Bibliografici

Arnold, K. E., Signals: Applying Academic Analytics. *Educause Quarterly*, 33,1, 2010,1-10.

Baker, R. S., and Paul Salvador Inventado. Educational data mining and learning analytics. *Learning Analytics*. Springer New York, 2014, 61-75.

Blondel V.D., Guillaume J., Lambiotte R., Lefebvre E., Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory and Experiment*. 10: P10008. <http://stacks.iop.org/1742-5468/2008/i=10/a=P10008.1oo> , 2008.

Chatti, M. A., Dyckhoff A.L., Schroeder U., Thüs H., A reference model for learning analytics. *International Journal of Technology Enhanced Learning (IJTEL)*, 4,5-6, 2012, 318-221.

Corposanto C., Lombi L., E-Methods and web society, Università Cattolica del Sacro Cuore, 2014.

Ferguson, R., Learning Analytics: fattori trainanti, sviluppi e sfide. TD tecnologie didattiche, 22, 3, 2014, 138-147.

Ferguson, R., Learning Analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4,5/6, 2012, 304-317.

Gutierrez-Santos S., S. Geraniou, E., Pearce-Lazard, D. & Poulouvassilis. A. (2012) Architectural Design of Teacher Assistance Tools in an Exploratory Learning Environment for Algebraic Generalisation. IEEE Transactions of Learning Technologies, 5 (4), 2012, 366-376.

Hadoop,<http://Hadoop.apache.org/2014>.

Koza J.R., Genetic programming: on the programming of computers by means of natural selection, MIT Press, Cambridge, MA, 1992.

Laney, Doug. 3D data management: Controlling data volume, velocity. Vol. 2. and variety. Technical report, META Group, 2001.

Manyika, J., Chui M., Bughin J., Brown B., Dobbs R. C., Roxburgh e Hung Byers A., Big Data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, 2011.

Manoochehri M., Data Just Right: Introduction to Large-Scale Data & Analytics. Addison-Wesley Professional, 2013.

Pappas C., <http://elearningindustry.com/Big-data-in-elearning-future-of-elearning-industry>, 2014.

Picciano, A. G.. The Evolution of Big Data and Learning Analytics in American Higher Education. Journal of Asynchronous Learning Networks,16, 3, 2012, 9-20.

Rezzani A., Big Data: Architettura, tecnologie e metodo per l'utilizzo di grandi basi di dati, Maggioli editore, 2013.

Reiss C., Tumanov A., Ganger G.R., Katz R.H., Kozuch M.A., Towards Understanding heterogeneous clouds at scale: Google Trace Analysis. Technical Report ISTC-CC-TR-12-101, Intel Science and technology center for Cloud computing, Carnegie Mellow University, Pittsburgh, PA,USA, 2012.

Siemens G., Gasevic D., Haythornthwaite C., Dawson S., Shum S.B., Ferguson R., Duval E., Verbert K., and Baker R. S. J. D. Open Learning Analytics: an integrated & modularized platform, 2011.

Slade S., Prinsloo P., Learning Analytics Ethical Issues and Dilemmas, American Behavioral Scientist 57,10, 2013, 1510-1529.

Snijders, C., Matzat, U., & Reips, U.-D. Big Data: Big gaps of knowledge in the field of Internet. International Journal of Internet Science, 7, 2012, 1-5.

Vaish G., Getting Start with NoSQL, Packt Publishing, 2013.

Wright A. H. Genetic Algorithms for Real Parameter Optimization, 1991.

Zedlewski, J., Sobti, S., Garg, N., Zheng, F., Krishnamurthy, A., Wang, R., YModeling Hard-Disk Power Consumption. FAST 3, 2003, 217-230.