

Percorso Autonomo Autorizzato

Title (Titolo)	Big Data Analytics (Analisi per dati ad alta complessità)
Chief (Referente responsabile)	(DMAT, PoliMi): prof. Anna Maria Paganoni
Supporting Coordinators (Altri referenti)	(DEIB, PoliMi): prof. Letizia Tanca; prof. Stefano Ceri (DMAT, PoliMi): prof. Luca Formaggia, prof. Piercesare Secchi (IBM): ing. Carla Milani, ing. Marco Monti
Scientific collaborations and partnerships (Collaborazioni scientifiche nazionali ed internazionali)	<ul style="list-style-type: none"> • EPFL École Polytechnique Fédérale de Lausanne, Switzerland • MRU – Cambridge, UK • École Centrale Paris – Supélec, France
Description and goals (Descrizione ed obiettivi)	<p>La società moderna è testimone di una sempre maggiore disponibilità di dati su piattaforme diverse - inclusi il web, blog e social network, le biblioteche digitali - e di forme diverse quali i profili temporali di click sui contenuti di una pagina web, o quelli delle chiamate da cellulare sulla rete di una data regione, le immagini satellitari per il remote sensing o quelle biomediche per la diagnostica, le cartelle cliniche e i dati testuali, i profili di campioni biologici e così via. Molti di questi dati non sono pienamente strutturati ma contengono informazioni preziose per sviluppare conoscenza sui fenomeni che regolano la società, come le opinioni emergenti nei social network, i temi prevalenti dei motori di ricerca, il comportamento di acquisto dei consumatori, la pratica clinica e epidemiologica, o la gestione e l'organizzazione di sistemi complessi. Le tradizionali tecniche di "knowledge discovery from data" non sono sempre adeguate all'uso dei big data, che per natura sono complessi e multidimensionali.</p> <p>Si ritiene quindi fondamentale creare una figura professionale con solide conoscenze informatiche, statistiche e gestionali in grado di progettare, controllare e condurre analisi basate su dati complessi.</p> <p>Il corso vuole porre le basi formative per i Data Scientist, figure professionali in grado di padroneggiare gli strumenti analitici di indagine come la statistica, la modellazione matematica e il data mining e, nello stesso tempo, far sviluppare capacità applicative e di problem solving.</p> <p>Le principali prerogative di un Data Scientist possono essere sintetizzate in:</p> <ul style="list-style-type: none"> ➤ competenze tecniche profonde in: statistica, informatica, trattamento dei dati strutturati e non strutturati, gestione della qualità dei dati e tecniche di visualizzazione; ➤ conoscenza delle tecnologie e delle piattaforme disponibili per la gestione ed il trattamento dei dati; ➤ conoscenze dei processi aziendali e del funzionamento delle organizzazioni; ➤ conoscenza delle problematiche e delle sfide del dominio settoriale (industria, finanza, settore pubblico, telecomunicazioni,...) di interesse.
Study Plan (Piano di studi)	<p>Il Piano di Studi si inserisce nel PSPA (Major) di <i>Applied Statistics</i>, introducendo ulteriori aspetti di informatica e di gestione di processo, e si realizza con la scelta dei seguenti corsi, oltre a quelli obbligatori nel Major:</p> <ol style="list-style-type: none"> 1. 095976 – <i>Advanced programming for scientific computing</i> (sem. II, 8 CFU): da selezionare "in effettivo" nella Tab. FREE; 2. 10 CFU di natura informatica, da selezionare "in autonomo" al posto del corso ING opzionale, di cui

	<p>5 CFU a scelta tra: 089202 – <i>Technologies for information systems</i> (ING-INF/05, sem. I, 5 CFU) 089183 – <i>Data bases 2</i> (ING-INF/05, sem. I, 5 CFU)</p> <p>e 5 CFU a scelta tra: 090950 – <i>Distributed systems</i> (ING-INF/05, sem. I, 5 CFU) 094743 – <i>Data management for the Web</i> (ING-INF/05, sem. I, 5 CFU) 095898 – <i>Computing infrastructures</i> (ING-INF/05, sem. II, 5 CFU)</p> <p>3. 10 CFU di natura gestionale, da selezionare “in autonomo” al posto del corso STAT del Major, a scelta tra: 096078 – <i>Accounting, finance & control</i> (ING-IND/35, sem. I, 10 CFU) 097380 – <i>Multichannel customer strategy</i> (ING-IND/35, sem. II, 10 CFU) 097370 – <i>Economics and management of multinational enterprises</i> (ING-IND/35, sem. I, 5 CFU) 097317 – <i>Management of design and innovation projects</i> (ING-IND/35, sem. II, 5 CFU)</p>
Past MSc theses (Alcune Tesi discusse)	<ul style="list-style-type: none"> • Nicholas Cazzaniga e Stefano D'Ettole (Ing. Mtm.), <i>Measuring reputational risk in electric utilities: a quantitative analysis of the stock market and Web reactions to negative events</i>, 2014 • Luca Ferrara (Ing. Mtm.), <i>Market Reactions to Reputational Events in the Oil&Gas Sector: Comparison between Companies 2014</i> • Rachele Biasi (Ing. Mtm.), <i>Uso delle misure di profondità per dati funzionali multivariati nella previsione di patologie: un'applicazione ai segnali elettrocardiografici</i>, 2013 • Ruben Binda (Ing. Mtm.), <i>A new approach to evaluate reputational loss: application to spread and oil and gas company</i>, 2013 • Mara Bernardi (Ing. Mtm.), <i>Registrazione di curve mediante l'algoritmo k-mean alignment ed estensione all'allineamento locale e di superfici</i>, 2013 • Claudia Belloni (Ing. Mtm.), <i>Sviluppo di modelli di stima e analisi del rischio reputazionale nel settore Oil&Gas</i>, 2013 • Emanuele Giani (Ing. Mtm.), <i>Metodi grafici ed inferenziali per l'identificazione di outliers: il caso del processo di cura di patologie cardiovascolari</i>, 2013 • Mirco Patriarca (Ing. Mtm.), <i>Development of the "fdakma" R package for the joint clustering and alignment of functional data with an application to neuronal spike trains data</i>, 2013 • Camilla Pezzotti (Ing. Mtm.), <i>Il processo di acquisto: studio del comportamento multicanale di un campione di consumatori</i>, 2013 • Teresa Pietrabissa (Ing. Mtm.), <i>Ricostruzione di hazard functions per lo studio della progressione di malattia in pazienti affetti da scompenso cardiaco</i>, 2013
(Available subjects for a MSc thesis) (Tesi disponibili)	<ul style="list-style-type: none"> • Analisi statistica per dati riferiti a varietà riemanniane bidimensionali, con applicazioni a dati di neuroimaging • Analisi statistica dei profili temporali di connettività di segnali EEG intracranici • Analisi statistica per object oriented data • Network analysis in healthcare and educational context • Analisi di database amministrativi ad alta dimensionalità
Internships (Tirocini)	Su richiesta dello studente è possibile attivare tirocini non curricolari con aziende italiane o straniere.
Job opportunities (Sbocchi lavorativi)	Aziende e società nel settore informatico, società di consulenza, società di progettazione e sviluppo software, i centri e le società che operano in ambito medico, biomedico e farmacologico, in ambito ecologico e nelle sezioni ricerca e sviluppo di grandi imprese, nel settore dei trasporti, delle telecomunicazioni, dell'energia.