

Business Intelligence per Big data



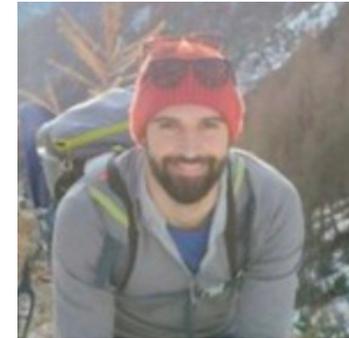
Politecnico
di Torino

Eliana Pastor
AA. 2023-2024
Politecnico di Torino

Eliana Pastor



Flavio Giobergia



Daniele Rege Cambrin



Eleonora Poeta



Organizzazione



- Inizio lezioni: 4/03/2024
- Fine lezioni: 11/06/2024

- Orario delle lezioni ed esercitazioni

Giorno	Orario	Aula	Tipo	
Lunedí	13:00-16:00	29	Lezione/Esercitazione	
Martedì	11:30-13:00	29	Lezione/Esercitazione	
Martedì	14:30-16:00	7I	Laboratorio	Squadra 2*
Martedì	16:00-17:30	7I	Laboratorio	Squadra 1*

Organizzazione Laboratorio



- I laboratori inizieranno a partire dalla terza settimana
- Laboratori a squadra
 - Divisione in squadra verra' comunicata a breve
 - Sovrapposizione con corso 'Gestione dell'innovazione e sviluppo prodotto ICT'

Giorno	Orario	Aula	Tipo	
Martedì	14:30-16:00	7I	Laboratorio	Squadra 2*
Martedì	16:00-17:30	7I	Laboratorio	Squadra 1*

- Business Intelligence: attività di gestione e analisi dei dati per il supporto delle decisioni aziendali
 - Data analytics
 - Data warehousing
 - Dashboard informative e tecniche di visualizzazione
 - No-SQL databases

- Business Intelligence: attività di gestione e analisi dei dati per il supporto delle decisioni aziendali
- Tecniche di analisi di dati
 - Tecniche ed algoritmi di data science e machine learning per trasformare la conoscenza a supporto del business
 - Tecniche di visualizzazione
- Studio delle basi di dati per il supporto decisionale (data warehouse)
- Architettura dei sistemi di data warehouse e delle metodologie di progettazione e sviluppo
 - OLAP (On Line Analytical Processing) dei dati
 - Introduzione alle basi di dati non relazionali (no-relational databases) a supporto dei Big Data
- Sviluppo di casi applicativi mediante strumenti di data science e data warehouse

- Business Intelligence: attività di gestione e analisi dei dati per il supporto delle decisioni aziendali
- Data analytics
 - Tecniche di preparazione dei dati
 - Tecniche di esplorazione dei dati
 - Processo di estrazione della conoscenza (data science pipeline)
 - Algoritmi di clustering ed indici di qualità
 - Algoritmi di estrazione di regole di associazione ed indici di qualità
 - Analisi di serie temporali e regressione
 - Algoritmi di classificazione e tecniche di validazione
 - Introduzione agli algoritmi di deep learning

- Data warehouse
 - architetture
 - progettazione e sviluppo
 - analisi OLAP (On Line Analytical Processing) dei dati
 - estensioni del linguaggio SQL per interrogare i data warehouse
 - Dashboard informative (basate su KPIs – Key Performance Indicators)
 - Tecniche di data visualization
- Basi di dati NO-SQL
- Tipi di basi di dati NO-SQL
- Concetti di base
- Esempi di interrogazioni

- Studio di casi applicativi di ampio interesse scientifico
 - Processo di estrazione della conoscenza mediante pipeline di data science
 - Dai dati alla conoscenza per trasformare il business (datadriven strategies & actions)
 - Presentazione di progetti di ricerca di data science
 - Progettazione di data warehouse
 - Analisi di requisiti per una corretta progettazione
 - Dashboard informative e infografiche
 - Interrogazioni di sistemi no relazionali

- Discussione di come progettare un workflow di data science
- Progettazione concettuale, logica e fisica di un data warehouse
- Interrogazione di un data warehouse mediante il linguaggio SQL esteso
- Interrogazione di un sistema NO-SQL mediante paradigma MapReduce

Esercitazioni di laboratorio in Aula



- Applicazioni di tecniche di data analytics
- Tool open source (i.e., RapidMiner)
- Interrogazione di una base di dati relazionale mediante il linguaggio SQL e le sue estensioni
 - DBMS: Oracle
- Progettazione di un data warehouse e analisi OLAP
 - Google Data Studio
 - Progettazione di un sistema NO-SQL e analisi
 - Tool open source: Mongo DB

- Progetto
 - Caso di studio da analizzare in gruppo (5 persone)
 - Consegna di una relazione scritta (con descrizione del contributo individuale)
 - Compilazione di 1-2 questionari
- Prova scritta
 - Progettazione concettuale, logica e fisica di un data warehouse
 - Interrogazione del data warehouse progettato mediante il linguaggio SQL e definizione di viste materializzate
 - 3 domande teoria/esercizi su data analytics
- Consegna di 2 homework
 - Data science
 - Basi di dati no relazionale

- Sito web del corso

dbdmg.polito.it/dbdmg_web/2024/business-intelligence-per-big-data-2023-2024/

Copia delle slide utilizzati a lezione

- Testi e risoluzioni di esercizi proposti in aula
- Testi e materiale utile per le esercitazioni in laboratorio
- Materiale utile per la risoluzione dei progetti
- Materiale integrativo
 - Modello relazionale e linguaggio SQL

- Data warehousing
 - Golfarelli, Rizzi. Data warehouse: teoria e pratica della progettazione, McGraw-Hill 2006
 - Kimball e altri, numerosi testi su metodologia e casi di studio, Wiley
- Data mining
 - Han, Kamber. Data mining: concepts and techniques, Morgan Kaufmann 2006
 - Tan, Steinbach, Kumar. Introduction to data mining, Pearson 2006
- Basi di dati no relazionali
 - K. Chodorow and M. Dirolf. MongoDB: the definitive guide. O'Reilly Media, 2010

- Data warehouse
 - <http://www.dwinfocenter.org>
 - <http://www.dwreview.com>
 - <http://kimballuniversity.com>
- Data mining
 - <http://www.kdnuggets.com>
- Tool open source
 - Pentaho: <http://www.pentaho.com/>
 - RapidMiner: <http://rapid-i.com/>
 - Mongo-DB: <https://www.mongodb.com/it>