

Academy Big Data

Corso di Alta Formazione

METODOLOGIE, TECNICHE E TOOL PER L'ANALISI DI BIG DATA

SITO WEB (sistema dolly accessibile agli iscritti)

<https://dolly.ingmo.unimore.it/2018/course/view.php?id=650>



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA



Calendario Lezioni e Laboratorio

	Venerdì – ore 14-18 Aula P1.4 (Fa-1d)	Sabato – ore 9-13 Laboratorio P2.6 (FA-2F LINFA)
Settimana I (18-19 Gennaio)	Big Data Technologies & Data Science Data & AI Ethics (Bergamaschi)	Introduzione a Python Architetture ad alte prestazioni (Fiameni)
Settimana II (25-26 Gennaio)	Big Data Integration GDPR & Data privacy (Bergamaschi)	Data integration & Entity Resolution: MOMIS & SparkER (Bergamaschi, Orsini)
Settimana III (8-9 Febbraio)	Data Analysis: metodi di Machine Learning Reti Neurali Ricorrenti (Turra, Pedrazzi, Morrone)	Implementazione R dei Metodi di ML (Turra, Pedrazzi)
Settimana IV (15-16 Febbraio)	Qualità dei Dati (DQ) nei database e nel Web Dimensioni DQ per i Big Data (Batini)	Modellazione dei processi Qualità nei processi aziendali (Pernici)
Settimana V (22-23 Febbraio)	Mining Massive Dataset: Associative Rules and LSH. (Guerra)	Metodi di progettazione del software Simulazione LEGO4SCRUM (Vincini)
Settimana VI (1-2 Febbraio)	Overview del processo di Business Intelligence Power BI (Beneventano)	Apache Spark - paradigma MapReduce MLlib – ML distribuito con Apache Spark (Gagliardelli)

Calendario Lezioni e Laboratorio

	Venerdì – ore 14-18	Sabato – ore 9-13	
Settimana I (18-19 Gennaio)	Big Data, Data Science, AI Ethics (Bergamaschi)	Python & High Performance Architectures (Fiameni)	LEZIONE
Settimana II (25-26 Gennaio)	Big Data Integration & GDPR (Bergamaschi)	Data Integration: MOMIS- SparkER (Bergamaschi, Orsini)	
Settimana III (8-9 Febbraio)	Data Analysis (Turra, Pedrazzi)	Data Analysis (Turra, Pedrazzi)	LABORATORIO
Settimana IV (15-16 Febbraio)	Data Quality (Batini)	Process Modelling & Management (Pernici)	
Settimana V (22-23 Febbraio)	Mining Massive Dataset: Association Rules & LSH (Guerra)	Software Engineering Methods LEGO4SCRUM (Vincini)	
Settimana VI (1-2 Marzo)	Business Intelligence (Beneventano)	Spark & MLIB (Gagliardelli)	

BIG DATA TECHNOLOGIES – DATA SCIENCE - DATA & AI ETHICS

- Introduzione
 - Introduzione ai Big Data
 - Machine Learning for Data Management & vice-versa
 - Data & AI Ethics
- Big Data: Principi e Sistemi
 - File system distribuiti (GFS), Consistenza Eventuale, Map reduce, teorema CAP
 - Google Table, Dynamo, Neo4J

Venerdì, 18 Gennaio

Prof. Sonia Bergamaschi

PYTHON - ARCHITETTURE AD ALTE PRESTAZIONI

- Introduzione a Python
 - Python come piattaforma per il calcolo scientifico
 - Esempi di codici

- Architetture ad alte prestazioni
 - Architettura di elaboratori
 - Cenni di calcolo parallelo
 - Evoluzione tecnologica

Sabato, 19 Gennaio

Dott. Giuseppe Fiameni

BIG DATA INTEGRATION - GDPR & DATA PRIVACY

- Big Data Integration
 - I Big Data nel contesto della Data Integration
 - Volume, Veridicità, Varietà, Velocità, Valore: nuovo approcci per l'integrazione dei Big Data

- GDPR & Data privacy
 - Big Data e GDPR The Regulation (EU) 2016/679

Venerdì, 25 Gennaio

Prof. Sonia Bergamaschi

DATA INTEGRATION & ENTITY RESOLUTION: MOMIS & SPARKER

- Data integration & Entity Resolution:
 - Metodi e processi per la Data Integration; introduzione del sistema MOMIS
 - Metodi e processi per l'Entity Resolution; introduzione del sistema SparkER
- MOMIS & SparkER
 - Demo del sistema MOMIS ed esempi di Data Integration
 - Demo del sistema SparkER ed esempi di Entity Resolution

Sabato, 26 Gennaio

Prof. Sonia Bergamaschi – Ing. Mirko Orsini

DATA ANALYSIS - RETI NEURALI RICORRENTI

- Data Analysis: metodi di Machine Learning
 - Apprendimento supervisionato (Decision Tree, Bagging, Boosting, Random Forest)
 - Apprendimento non supervisionato (K-means, dbScan, GMM,...)
- Reti Neurali Ricorrenti
 - Introduzione alle RNN e varianti (LSTM e GRU)
 - Dimostrazione pratica di una semplice applicazione con Keras / Tensorflow.

Venerdì, 8 Febbraio

Dott.ssa Roberta Turra/Dott. Giorgio Pedrazzi – Ing. Giovanni Morrone

IMPLEMENTAZIONE R DEI METODI DI ML

- Costruzione di applicazioni di Machine Learning
 - Interfaccia Rstudio
 - R base (oggetti, vettori, matrici, dataframe, lettura dati)
 - Installazione pacchetti
 - R parallelo (cenni)

- Pacchetti R specifici
 - Clustering con kmeans, dbscan e mclust
 - Classificazione con rpart e randomforest

Sabato, 9 Febbraio

Dott.ssa Roberta Turra/Dott. Giorgio Pedrazzi

QUALITÀ DEI DATI (DQ) NEI DATABASE E NEL WEB

- Dimensioni della qualità dei dati in basi di dati
- La qualità dei dati in uso
- Qualità dei dati nel Web: veridicità delle fonti e dimensioni DQ per i big data

Venerdì 15 Febbraio

Prof. Carlo Batini

MODELLAZIONE DEI PROCESSI E QUALITÀ DEI PROCESSI

- Concetti di modellazione del processo
- BPMN: notazione di modellazione dei processi aziendali
- Qualità nei processi aziendali

Sabato, 16 Febbraio

Prof. Barbara Pernici

MINING MASSIVE DATASET: ASSOCIATION RULES & LSH

- Mining massive datasets:
 - Market basket analysis
 - Locality-sensitive hashing per trovare documenti simili.

Venerdì, 22 Febbraio

Prof. Francesco Guerra

METODI DI PROGETTAZIONE DI UN PROGETTO SOFTWARE

- Metodi di progettazione del software
 - Il Processo Unificato (UP)
 - Metodi agili di sviluppo del software e loro principali caratteristiche
 - SCRUM e DevOps

- LEGO4SCRUM
 - Gioco di simulazione metodologia SCRUM attraverso i LEGO

Sabato, 23 Febbraio

Prof. Maurizio Vincini

INTRODUZIONE ALLA BUSINESS INTELLIGENCE E USO DEL TOOL POWER BI

- Overview del processo di Business Intelligence
 - Analisi Multidimensionale e sistemi OLAP
 - Architetture per la Business Intelligence

- Strumenti BI per la visualizzazione interattiva dei dati
 - Power BI
 - Power Pivot, Power Query and Power View

Venerdì, 1 Marzo

Prof. Domenico Beneventano

APACHE SPARK & MLlib

- Apache Spark
 - Introduzione al paradigma MapReduce
 - Introduzione al framework Apache Spark
 - Primitive di Spark

- MLlib
 - Machine Learning distribuito con Apache Spark
 - Esempi in linguaggio Python per provare le potenzialità di Spark

Sabato, 2 Marzo

Ing. Luca Gagliardelli