

SUPSI

CAS Big Data Analytics e Machine Learning

Descrizione

Codice

DTI DA

Presentazione

Nel corso degli ultimi decenni i dati sono diventati sempre più un patrimonio fondamentale per qualunque organizzazione. Essi rappresentano sia una sfida che un'opportunità: alla complessità della gestione di volumi di dati sempre più grandi si affianca la possibilità di utilizzarli per prendere decisioni che possano costituire un vantaggio competitivo per l'organizzazione o per aprire nuove opportunità di business. Inoltre, col crescere della diffusione di sistemi intelligenti integrati nell'Internet of Things, è sempre più necessario che siano i sistemi stessi a prendere decisioni basandosi sui dati a disposizione e sulla conoscenza che incorporano. L'utilizzo intelligente dei dati è quindi diventato una necessità nel mondo di oggi e del futuro ed è un elemento fondamentale della cosiddetta rivoluzione digitale.

I sistemi, aziendali o piattaforme internet, che concentrano un'enorme ed eterogenea mole di dati, i cosiddetti Big Data, sono l'ambiente ideale in cui far operare algoritmi di data mining, al fine di estrarre conoscenza dai dati e supportare le applicazioni di profiling, raccomandazione, decisione, e non solo.

Obiettivi

- Conoscere le tecnologie e le piattaforme utilizzabili per l'analisi di Big Data
- Conoscere e imparare a utilizzare algoritmi per data analysis
- Capire e implementare modelli di classificazione, regressione, clustering e raccomandazione; usarli per analizzare grandi quantità di dati
- Conoscere il Machine Learning e imparare ad applicarlo a dei dataset
- Conoscere strumenti e piattaforme software con cui realizzare applicazioni di data analysis.

Destinatari

Professionisti che operano nel campo della gestione dei dati in azienda e che intendono utilizzare i dati nell'ambito della business intelligence o nella predictive analytics.

Requisiti

Conoscenze di base di gestione e di analisi dei dati e di programmazione. Nell'ambito di alcuni corsi verrà utilizzato il linguaggio Python, a chi non ne avrà le basi verrà indicato un tutorial da seguire prima dell'inizio del corso.

Certificato

Certificate of Advanced Studies SUPSI

Crediti di studio

10 ECTS

Contenuti

Programma

Modulo 1: Machine Learning e Data Mining (36 ore lezione)

- Introduzione al machine learning/data mining.
- Uso grafico di Weka.
- Preparazione dei dati (discretizzazione, pulizia, dati mancanti, selezione delle variabili).
- Apprendimento supervisionato (classificatori principali e meta-classificatori: bagging, boosting).
- Valutazione empirica dei classificatori.
- Dati temporali (serie storiche).
- Cenni a bias-variance decomposition, classificazione cost-sensitive, test statistici di confronto.
- Introduzione alle reti Bayesiane.
- Grafi causali, d-separazione, probabilità condizionata.
- Uso di Samlam.
- Esempi di casi pratici durante tutto il modulo.

Modulo 2: Tecnologie per i Big Data (28 ore-lezione)

- Hadoop Distributed File System (HDFS)
- Hadoop MapReduce
- Apache Spark
- Spark cluster manager
- Stream e event processing
- Data visualisation Tools
- Esercizi

Modulo 3: Big Data analytics con Spark (24 ore-lezione)

- Classificazione: descrizione degli algoritmi (modelli lineari, naive Bayes, alberi di decisione) e dell'implementazione Spark.
- Regressione: descrizione degli algoritmi (regressione lineare, random forest) e dell'implementazione Spark.
- Clustering: descrizione degli algoritmi (k-means, mistura di Gaussiane) e caso di studio Spark.
- Raccomandazione: descrizione degli algoritmi (collaborative filtering) e dell'implementazione in Spark.
- Caso di studio pratico: raccomandazione di film, analisi di dati di bike sharing, sentiment analysis.

Modulo 4: Piattaforme cloud per Analytics e Machine Learning

(12 ore-lezione)

Studio e utilizzo delle principali piattaforme cloud per Analytics e Machine Learning :

- Amazon Machine learning
- Google Prediction API
- Microsoft Machine Learning Studio
- open source: cloudera, hortonworks
- studio di caso

Modulo 5: Progetto finale

Il progetto finale consentirà allo studente di mettere in pratica le competenze acquisite nei moduli precedenti ed entrerà a far parte della valutazione finale dello studio. Il tema del progetto verrà proposto da uno dei docenti del corso e verrà svolto dallo studente autonomamente con interazioni e incontri periodici con il relatore.

Durata

100 ore-lezione più progetto personale

Responsabile/i

Roberto Mastropietro, Direttore, Istituto sistemi informativi e networking

Relatore/i

Patrick Ceppi, docente-ricercatore, Istituto Sistemi Informativi e Networking, SUPSI

David Huber, Software Engineer & Researcher, Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, SUPSI

Informazioni

Iscrizione ai corsi

Entro il 1. marzo 2019

Date

Modulo 1: 11, 13, 20, 25, 27 marzo, 1, 3, 8, 10 aprile 2019

Modulo 2: 15,17, 29 aprile, 6, 8, 13, 15 maggio 2019

Modulo 3: 20, 22, 27, 29 maggio, 3, 5 giugno 2019

Modulo 4: 11, 18, 25 settembre 2019

Modulo 5: 2, 9, 16, 23 ottobre, 6, 13 novembre 2019

La presentazione del progetto, relativa al modulo 5, avrà luogo nel corso dell'anno accademico 2019-2020.

Orari

17.30-21.00

Luogo

SUPSI, Dipartimento tecnologie innovative, Manno

Costo

CHF 4'800.–

Tali costi comprendono il rilascio dei certificati e la documentazione didattica

È previsto uno sconto del 10% per i membri individuali dell'Associazione Ticinese Elaborazione Dati (ATED) e della Swiss Engineering (ATS).

Informazioni

SUPSI, Dipartimento tecnologie innovative

Galleria 2

CH-6928 Manno

T +41 (0)58 666 66 84

F +41 (0)58 666 65 71

dti.fc@supsi.ch

www.supsi.ch/dti