

SHORT MASTER

BIG DATA ANALYTICS PER DATI TABELLARI E TESTUALI

CORSO INTRODUTTIVO
TEORICO PRATICO

Edizione 2024

2 GIORNATE
14 ORE

In collaborazione con



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

Dipartimento di
Ingegneria "Enzo Ferrari"



RETE ALTA TECNOLOGIA
EMILIA-ROMAGNA
HIGH TECHNOLOGY NETWORK
TECNOPOLO MODENA

BIG DATA ANALYTICS PER DATI TABELLARI E TESTUALI - Corso introduttivo teorico pratico

Il concetto di Big Data si riferisce all'insieme di tecnologie e pratiche utilizzate per raccogliere, archiviare, analizzare e interpretare grandi volumi di dati, che sono troppo complessi o troppo vasti per essere gestiti con strumenti tradizionali. Questi dati possono provenire da diverse fonti, come social media, sensori IoT, transazioni online, e sono caratterizzati dalle cosiddette "3V" (Volume, Varietà, Velocità), a cui spesso se ne aggiungono altre come Veridicità e Valore. Con il concetto Big Data Analytics ci si riferisce al processo tramite cui si esaminano e analizzano grandi volumi di dati per scoprire pattern nascosti, correlazioni sconosciute, tendenze di mercato, preferenze dei clienti e altre informazioni utili che possono aiutare le organizzazioni a prendere decisioni informate. Questa disciplina combina tecniche avanzate di analisi con l'uso di tecnologie di elaborazione dei dati di ultima generazione per estrarre valore da dati complessi e di grandi dimensioni. In sintesi, Big Data Analytics è una disciplina essenziale per sfruttare al massimo il potenziale dei dati che vengono raccolti. L'uso efficace di queste tecniche può portare a una migliore comprensione del mercato, innovazione e competitività, ma richiede anche l'adozione di tecnologie avanzate e la gestione delle sfide legate alla sicurezza e alla qualità dei dati.

Con queste premesse, Fondazione Democenter in collaborazione con l'Università di Modena e Reggio Emilia, propone un corso di 14 ore totali in presenza, con l'obiettivo di fornire una panoramica tecnica sulle principali metodologie di **data mining** e **machine learning per l'analisi e l'estrazione di conoscenza** da dati strutturati tabellari e non strutturati testuali.

Nella prima parte, il corso si concentra sull'**analisi dei dati strutturati** (per esempio le tipiche tabelle numeriche con dati transazionali o rilevazione di sensori), utilizzando algoritmi classici di data mining e machine learning con l'ausilio di librerie Python come scikit-learn, pandas e numpy. La seconda parte approfondisce l'**analisi dei dati testuali** (per esempio commenti e descrizione di prodotti), introducendo embeddings per la rappresentazione numerica del testo e tecniche di NLP per l'integrazione dei dati testuali nei modelli di machine learning, sfruttando librerie dell'ecosistema Hugging Face che implementano architetture basate su transformers.

I partecipanti acquisiranno competenze teoriche e pratiche* fondamentali per l'applicazione di tecniche di machine learning e data mining a problemi reali di analisi dati in ambito aziendale.



GIOVEDÌ 7 E 14 NOVEMBRE 2024



9:00 – 17:00



TECNOPOLO DI MODENA - Via P. Vivarelli 2, 41125 Modena

***Nelle lezioni di laboratorio (pratica) i partecipanti lavoreranno con il proprio computer su cui potranno installare software e applicazioni necessarie per lo svolgimento delle attività.**

DESTINATARI

Responsabili e tecnici IT coinvolti nei processi di digitalizzazione aziendale, integratori di sistemi ,architetti software, sviluppatori, ingegneri di processo.

Pre-requisito: è richiesta una conoscenza di base del linguaggio Python.



OBIETTIVI

- Conoscere il flusso completo dell'analisi dei dati, dal processamento iniziale, all'allenamento e successiva valutazione di modelli.
- Analizzare ed estrarre conoscenza da dati tabellari e testuali utilizzando tecniche di machine learning.
- Sviluppare competenze di base sull'applicazione di algoritmi di machine learning per classificazione, regressione e clustering.
- Utilizzare metodi di Natural Language Processing (NLP) per il trattamento e l'analisi dei dati testuali.

DOCENTI

Prof. Ing. Francesco Guerra, Professore Ordinario Dipartimento di Ingegneria "Enzo Ferrari", Università degli Studi di Modena e Reggio Emilia.

Dott. Matteo Paganelli, Assegnista di ricerca presso il Dipartimento di Ingegneria "Enzo Ferrari", Università degli Studi di Modena e Reggio Emilia.

7 NOVEMBRE 2024, 9.00 - 17.00

BIG DATA ANALYTICS PER DATI TABELLARI

Teoria (3 ore)

- La pipeline dell'analisi dei dati: il flusso completo dell'analisi dei dati, dalla raccolta e pulizia delle informazioni alla modellazione, valutazione e interpretazione dei risultati.
- Tecniche di Machine Learning per l'analisi dei dati: panoramica delle principali, tecniche di Machine Learning utilizzate per analizzare grandi quantità di dati, incluse tecniche di classificazione, regressione, regole associative e clustering.
- Criticità: il Machine Learning in produzione, il bias, explanation: le sfide legate alla messa in produzione dei modelli di Machine Learning, con particolare attenzione a problemi di bias, equità e spiegabilità delle decisioni algoritmiche.

Applicazioni (4 ore)

- Analisi ed esplorazione dei dati in Python: identificare valori nulli e outliers, analizzare la distribuzione delle caratteristiche dei dati e la loro correlazione in maniera visuale.
- Pulizia e trasformazione dei dati in Python: pulire i dati da feature non necessarie, da valori nulli e duplicati. Normalizzare i valori numerici e codificare valori testuali/categorici.
- Creare modelli di machine learning con Sklearn: realizzare una standard pipeline di training e valutazione di modelli di machine learning. Risolvere scenari di classificazione, regressione e clustering.

14 NOVEMBRE 2024, 9.00 - 17.00

BIG DATA ANALYTICS PER DATI TESTUALI

Teoria (3 ore)

- Natural Language Processing (NLP): overview.
- Introduzione al text retrieval.
- Rappresentazione di testo attraverso embedding.
- Language model.

Applicazioni (4 ore)

- Codifica di testi con embedding in Python: Utilizzo di embedding da modelli pre-trainati basati su transformer.
- Guida allo sviluppo di modelli NLP in Python: Allenamento di modelli basati su transformers per risolvere svariati NLP task.

BIG DATA ANALYTICS PER DATI TABELLARI E TESTUALI - Corso introduttivo teorico pratico

DATI DI ISCRIZIONE DEL PARTECIPANTE

Cognome e nome
Telefono.....Cell.E-mail.....
Titolo di studio
Ruolo ricoperto all'interno dell'azienda

AZIENDA DI APPARTENENZA (in caso di partecipazione a titolo aziendale)

Ragione socialePartita I.V.A./C.F.....Codice SDI.....
Attività dell'azienda
Indirizzo.....CAP.....Comune.....Prov.
N. dipendentiTelFaxE-mail

DATI PER LA FATTURAZIONE (compilare solo se differenti rispetto ai dati dell'azienda)

Intestazione e indirizzo
Partita I.V.A./ C.F.....Codice SDI

QUOTA DI PARTECIPAZIONE (barrare la scelta)

Prezzo intero

720 € + Iva

Prezzo Soci Democenter

645 € + Iva

Prezzo secondo iscritto*

720€ 648 € + Iva

Prezzo secondo iscritto Soci Democenter*

645€ 580 € + Iva

***Si applica uno sconto del 10% a partire dal 2°iscritto.**

MODALITÀ DI ISCRIZIONE

L'iscrizione dovrà avvenire entro il 3° giorno lavorativo antecedente l'inizio del corso. L'iniziativa verrà realizzata al raggiungimento del numero minimo di 8 iscritti. In caso di mancato raggiungimento di tale numero, Fondazione Democenter-Sipe si riserva la facoltà di disdire il corso, comunicandolo all'indirizzo del partecipante entro 2 giorni dalla data di inizio prevista. In tal caso, al partecipante /Azienda che ha già provveduto al pagamento della quota di iscrizione verrà offerta la possibilità di partecipare ad un altro corso o verrà restituita la quota di iscrizione.

La presente scheda dovrà essere inviata alla Fondazione Democenter-Sipe via email a formazione@fondazionedemocenter.it. Per chiarimenti è possibile contattare la Fondazione Democenter allo 059 2058153.

CONDIZIONI DI PAGAMENTO

La quota di iscrizione deve essere versata al momento della conferma del corso. Il pagamento deve essere effettuato mediante bonifico Bancario intestata a **Fondazione Democenter-Sipe codice IBAN: IT44C053871290500000551764 presso Banca Popolare dell'Emilia - Romagna, Ag. 5 di Modena**. Fondazione Democenter-Sipe provvederà all'invio della fattura elettronica al ricevimento della quota di iscrizione.

DISDETTA DELLA PARTECIPAZIONE

Qualsiasi rinuncia deve pervenire, in forma scritta, entro 4 giorni lavorativi dall'inizio del corso. In caso di rinuncia pervenuta dopo tale termine o di mancata presenza del partecipante ad inizio corso o di ritiro durante lo stesso Fondazione Democenter è autorizzata a trattenere l'intera quota se già versata.

I dati raccolti saranno trattati ai sensi del regolamento europeo sulla protezione dei dati (Reg. UE 2016/679).

Per maggiori informazioni sul trattamento, sulla privacy e sui diritti esercitabili vedi anche l'informativa sul sito www.democentersipe.it/privacy/

Si fornisce il consenso al trattamento dei propri dati personali in riferimento all'informativa ricevuta

SÌ NO

Data

Timbro e firma