UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Laboratorio 2 - Big data analytics: modelli di machine learning e GenAI per l’analisi dei dati

2425-2-F8802N074

Obiettivi formativi

Comprendere le sfide e le opportunità dei Big Data; Acquisire conoscenze su normative europee e strumenti riguardanti la raccolta e l'uso di tracce digitali; Utilizzare modelli di machine learning per l’analisi di testi; Integrare l’Intelligenza Artificiale nella ricerca sociale.

Contenuti sintetici

Definizione ed utilizzi dei Big Data; GDPR e DSA applicati alla raccolta dati digitali nel contesto Europeo; Introduzione a BERT e sue applicazioni nell'analisi testuale; Introduzione alle API di IA e loro applicazioni in diversi compiti della ricerca sociale.

Programma esteso

Lezione 1, 18 Novembre 2024, 14.30-18.30 (LAB714 - U7): Cosa sono e a cosa servono i Big Data
Argomenti principali:

- Definizione e caratteristiche dei Big Data (volume, varietà, velocità, veridicità, valore).
- Opportunità dei Big Data in diversi settori (ricerca sociale, marketing, automazione).
- Sfide associate ai Big Data: gestione, qualità e privacy.
- Le tracce digitali come principale fonte di Big Data nella ricerca sociale.
  Attività:
- Discussione di casi reali di Big Data in vari settori.

- Brainstorming su potenziali usi e problemi dei Big Data.
- Panoramica sui tool "reday-to-use" per l'analisi real time delle tracce digitali

Lezione 2, 25 Novembre 2024, 14.30-18.30 (LAB719 - U7): Raccolta dati e il Digital Services Act
Argomenti principali:

- Regolamentazioni per l'accesso e l'utilizzo di dati nel contesto europeo: dalla General Data Protection Regulation (GDPR) al Digital Services Act (DSA).
- Modalità di raccolta delle tracce digitali: fonti e strumenti.
  Attività:
- Analisi di come il DSA influenzi l'accesso a diverse tipologie di dati.
- Esercizi di raccolta delle tracce digitali tramite Zeeschuimer e tramite Hyphe.

Lezione 3, 2 Dicembre 2024, 14.30-18.30 (LAB719 - U7): Introduzione a Python
Argomenti principali:

- Introduzione all'utilizzo di Python per il data wrangling (numpy, pandas, matplotlib).
- Introduzione all'ambiente cloud CoLab per la programmazione condivisa.
  Attività:
- Esercizi di pre-processing, pulizia e visualizzazione dati con Python.
- Esplorazione guidata del dataset Twitter (circa 2 milioni di tweets) sulle alluvioni in Emilia Romagna.

Lezione 4, 9 Dicembre 2024, 14.30-18.30 (LAB719 - U7): Il Machine Learning per l'analisi di testi
Argomenti principali:

- Architettura di BERT (Bidirectional Encoder Representations from Transformers) e sue applicazioni nell'elaborazione del linguaggio naturale (NLP).
- Vantaggi dell'open source rispetto a prodotti commerciali chiusi.
  Attività:
- Ricognizione e presentazione degli applicativi online basati su BERT come ClimateBERT.
- Esercizi di classificazione del testo e sentiment analysis con BERT.

Lezione 5, 16 Dicembre 2024, 9.30-13.30 (LAB712 - U7): Utilizzare le API di Intelligenza Artificiale
Argomenti principali:

- Introduzione alle API di Intelligenza Artificiale: funzionamento e potenzialità.
- Panoramica su varie API di Intelligenza Artificiale tramite Replicate.
  Attività:
- Comprendere e usare una documentazione API.
- Esercizi di trascrizione di interviste con WhisperAI e classificazione di immagini tramite Memespector o ImageSorter o PixelPlot.

Lezione 6, 16 Dicembre 2024, 14.30-18.30 (LAB712 - U7): Seminario e progetto finale
Argomenti principali:

- Seminario del prof. Guido Anselmi a partire dal suo ultimo libro "Aprire la Scatola Nera. Corso riflessivo di Sociologia Computazionale"
- Discussione con l'autore sulle implicazioni dei Big Data Analytics nel capitalismo di piattaforma (ad esempio il caso Airbnb).
- Revisione degli argomenti trattati nel corso del laboratorio.
- Implementazione di un progetto finale in piccoli gruppi in stile hackathon.
  Attività:
- Lavoro di gruppo per la progettazione e realizzazione del progetto finale.
- Presentazione del progetto agli altri partecipanti e valutazione finale.

Prerequisiti

Concetti basi di programmazione e logica della programmazione.

Metodi didattici

Il corso è progettato come un'esperienza pratica di apprendimento, basata sul principio del 'learning by doing'. Il metodo di insegnamento sarà quello della 'didattica capovolta': le letture devono essere preparate in modo autonomo prima delle lezioni, mentre in aula saranno presentati concetti chiave e veranno svolte esercitazioni guidate.

Modalità di verifica dell'apprendimento

La valutazione sarà basata sulla qualità del progetto finale e la partecipazione attiva durante le lezioni.

Testi di riferimento

Dispensa fornita dal docente ad inizio corso.

Sustainable Development Goals

ISTRUZIONE DI QUALITÁ