

Social Media Data Analysis 2023/2024

Introduzione al Corso

Francesco Ragusa

francesco.ragusa@unict.it

<https://iplab.dmi.unict.it/ragusa/>

<https://iplab.dmi.unict.it/fpv/>



Informazioni Generali

- Nome del corso: Social Media Data Analysis
- Docente: Francesco Ragusa
- CFU: 6
- Orario delle lezioni: Mercoledì e Venerdì 17.00-19.00
- Luogo delle lezioni: Aula 22
- Materiale didattico e annunci: su Microsoft Teams (codice vo3q1m6)
- Ricevimento: Lunedì 14.30-16.30

Social Media Data Analysis

- Il corso si propone di approfondire teorie e tecniche per l'analisi di dati multimediali (testo, immagini, tag, metadati, ecc...) presenti nei Social Media;
- Principali moduli:
 - Social media: definizione, caratteristiche e breve storia;
 - Estrapolazione e trattamento dei dati presenti sul Social Media;
 - Algoritmi di Machine Learning applicati al contesto dei Social Media;
 - Metodi per la rappresentazione e il processamento di testi;
 - Sistemi di raccomandazione e sentiment analysis;
 - Algoritmi per il trattamento delle immagini;
- Contenuti e Syllabus:
<https://web.dmi.unict.it/corsi/I-31/insegnamenti?seuid=544C5A47-37C7-49BC-AD4E-96427883DE38>

Tipologia delle lezioni

Tre tipologie di lezioni:

- Lezioni frontali:
 - Discuteranno gli aspetti teorici delle tecniche affrontate nel corso, con esempi applicativi. Svolte in aula mediante slide e lavagna (digitale). Pensate per favorire l'interazione tra gli studenti e il docente.



Tipologia delle lezioni

Tre tipologie di lezioni:

- Laboratori:
 - Svolti in aula mediante Python e Jupyter Notebooks. Ogni studente verrà incoraggiato a portare il proprio laptop in aula. Hanno lo scopo di permettere di toccare con mano librerie e strumenti utili all'applicazione dei concetti visti a lezione;



Tipologia delle lezioni

Tre tipologie di lezioni:

- Seminari accademici e industriali:
 - Svolti da ospiti esterni che mostreranno come le tecniche viste durante il corso possono essere utilizzate in contesti accademici e industriali. Hanno lo scopo di fornire agli studenti una visione ampia dell'impatto delle competenze acquisite a lezione sulla vita reale.



Materiale Didattico

- Il docente distribuirà mediante Microsoft Teams le slide delle lezioni (corredate di note) e gli appunti delle sessioni di laboratorio. Il materiale didattico è prevalentemente in inglese;
- Le note distribuite dal docente costituiscono una traccia degli argomenti svolti a lezione e contenuti nel programma e vanno integrate mediante gli appunti presi a lezione e le letture consigliate di volta in volta;
- Una lista di libri consigliati è disponibile sul syllabus (<https://web.dmi.unict.it/corsi/l-31/insegnamenti?seuid=544C5A47-37C7-49BC-AD4E-96427883DE38>);
- Su richiesta, il docente potrà consigliare ulteriori testi sui quali approfondire gli argomenti visti a lezione.

Testi di riferimento

- R. Zafarani, M. A. Abbasi, H. Liu, [Social Media Mining - An Introduction](#), Cambridge University Press, 2014
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). [An introduction to statistical learning](#). New York: springer.
- C. Bishop, [Pattern Recognition and Machine Learning](#), Springer, 2006
- E. Alpaydin, [Introduction to Machine Learning](#), The MIT Press, 2009
- Y. Bengio, I. J. Goodfellow, A. Courville, [Deep Learning](#), Book in preparation for MIT Press, 2015
- Duda, P. E. Hart, D. G. Stork, [Pattern Classification \(2nd ed.\)](#), Wiley, 2000
- R. Szeliski, [Computer Vision: Algorithms and Application](#), Springer 2010
- J. Leskovec, A. Rajaraman, J. D. Ullman, [Mining of Massive Datasets](#), Cambridge University Press, 2011

Modalità D'esame

L'esame consiste in tre parti:

1. Uno scritto sugli argomenti del corso (peso sul voto finale: 60%);
2. Un "tutorial" in cui lo studente descrive le API di un Social Media a scelta, mostra come usarle per estrarre dei dati, e riporta un esempio di analisi di tali dati per estrarre conoscenza dal social media (peso sul voto finale: 20%);
3. Un progetto in cui lo studente deve costruire un software capace di analizzare i dati presenti su un social media per risolvere un problema di carattere pratico (peso sul voto finale: 20%);

Modalità D'esame/Scritto

Lo scopo dello scritto è la verifica dell'apprendimento dei concetti teorici e tecnici visti a lezione.

- Lo scritto verrà svolto nella data ufficiale dell'appello;
- Lo scritto consiste in 3 domande a risposta aperta alle quali lo studente risponderà in forma scritta in 2 ore. Esempi di domande sono disponibili sul syllabus:

<https://web.dmi.unict.it/corsi/l-31/insegnamenti?seuid=544C5A47-37C7-49BC-AD4E-96427883DE38>;

- Il voto finale dello scritto sarà in trentesimi.

Modalità D'esame/Tutorial

Lo scopo del tutorial è la verifica dell'apprendimento delle competenze relative all'estrazione e manipolazione dei dati dai social media.

- Il social media scelto va concordato brevemente con il docente;
- Il tutorial va svolto mediante Jupyter notebook in Python;
- Il tutorial deve coprire le seguenti sezioni:
 - Introduzione: breve introduzione al social media scelto per il tutorial;
 - Accesso alle API: descrizione del processo di accesso alle API (registrazione al sito, API key, API secret);
 - API REST ufficiale: descrizione dell'uso delle API ufficiali REST, riportando almeno un esempio di utilizzo;
 - [opzionale] API di terze parti: descrizione dell'uso di API di terze parti (es. wrapper in Python);
 - Estrazione dati e analisi: utilizzo delle API per estrarre dati dal social media e analizzarli per estrarre informazione dal social media. L'analisi può essere anche molto semplice;
 - Conclusioni: breve paragrafo sulle conclusioni tratte dal tutorial.
- La consegna del tutorial avviene via mail. Lo studente dovrà allegare il notebook jupyter e una versione PDF del tutorial.

Modalità D'esame/Progetto

- Lo scopo del progetto è la verifica delle competenze pratiche di estrazione e analisi di dati provenienti da social media mediante agli algoritmi visti a lezione;
- Il progetto deve essere proposto dallo studente e concordato con il docente, che fornirà spunti sul come rendere il progetto completo;
- Il progetto deve includere una relazione (scritta in Jupyter, word o in latex – dimensione font 12) e comprendere le seguenti sezioni:
 - Introduzione: definisce il problema affrontato e discute perché è interessante;
 - Dati: descrive i dati utilizzati durante il progetto. Nel caso in cui i dati siano stati acquisiti dallo studente, descrive come questi dati sono stati acquisiti;
 - Metodo proposto: descrive gli algoritmi implementati per la risoluzione del problema;
 - Risultati: discute i risultati ottenuti dall'algoritmo considerato, riportando una analisi quantitativa o qualitativa;
 - Conclusioni: riporta le conclusioni tratte dal progetto.
- La consegna deve avvenire via mail. Lo studente deve fornire un archivio contenente i dati utilizzati, il codice, e la relazione in formato PDF.

Modalità D'esame/Tempi

- L'esame scritto deve essere sostenuto nella data di esame ufficiale dell'appello. Gli studenti iscritti al CINAP possono richiedere preventivamente di sostenere un orale, qualora lo preferissero.
- Il progetto e il tutorial vanno preferenzialmente consegnati durante le sessioni di esame, ma possono essere consegnate via email in qualsiasi momento dell'anno;
- La verbalizzazione avverrà quando i tre elaborati verranno ricevuti e valutati;

Modalità D'esame/Tempi II

Alcune note sugli appelli:

- Generalmente è concesso a tutti gli studenti di partecipare a tutti gli appelli disponibili (compresi quelli riservati ai lavoratori e ai non in corso). Qualora non fosse per voi possibile prenotarvi a un appello, vi chiedo di contattarmi;
- È possibile sostenere lo scritto in una sessione e avere l'esame finale verbalizzato in un'altra sessione. Di tanto in tanto, i vecchi appelli vengono chiusi. In tal caso, è necessario prenotarsi per un nuovo appello nel quale non sosterrete lo scritto, ma vi verrà confermato il voto precedente;
- Il docente terrà nota dei voti degli scritti fino alla fine dell'erogazione del corso successivo. In pratica, una volta sostenuto lo scritto, avete tempo fino al Dicembre successivo per completare l'esame con progetto e tutorial.

Comunicazione

- Le comunicazioni ufficiali sul corso da parte del docente avverranno mediante Microsoft Teams;
- Gli studenti possono comunicare con il docente mandando una email all'indirizzo francesco.ragusa@unict.it. **Si prega di non contattare il docente mediante altri mezzi quali Microsoft Teams, Telegram, Facebook, ecc.;**
- Alcune semplici regole da seguire quando si spedisce una email:
 - Utilizzare l'indirizzo studium fornito dall'Ateneo. Questo serve a minimizzare la probabilità che la vostra mail finisca nello spam;
 - Firmarsi sempre alla fine della mail, in modo che possa sapere chi mi sta contattando;
 - Identificarsi sempre in relazione al corso (es. "sono uno studente di Social Media Data Analysis 2023/2024");
 - Rileggere la mail prima di spedirla.

