

# Progetti IA

June 3, 2020

## 1 Idee progettuali

Le idee progettuali di cui sotto non devono essere sviluppati fino alla realizzazione effettiva ma, come detto a lezione, devono essere l'argomento per un documento nel quale si presenti e sviluppi l'idea.

L'informatica e l'intelligenza artificiale possono dare un contributo sostanzioso alla ricerca in ambito medico e quindi anche alla scoperta di soluzioni e cure per il COVID-19. Sotto alcuni temi:

1. Metodi predittivi basati su tecniche di apprendimento automatico sono già ampiamente usati in molti settori medici, dall'oncologia alla genetica. Nel caso della pandemia attuale questi metodi possono essere usati in molti contesti diversi quali, ad esempio:
  - (a) Valutazione dei sintomi per individuare le priorità nel triage;
  - (b) Analisi dei dati della terapia intensiva per la valutazione delle diverse opzioni terapeutiche disponibili;
  - (c) Analisi dei dati epidemiologici per il controllo della diffusione del virus.
2. Metodi della bioinformatica possono essere usati per l'analisi di dati molecolari e proteici allo scopo di ridurre la complessità nell'ambito della ricerca di nuovi farmaci per il virus.
3. Metodi di Natural Language Processing possono essere usati per facilitare la ricerca di informazioni nella letteratura scientifica sul COVID-19 (un dataset, contenete già 26000 articolo scientifici in questo ambito è già stato creato).
4. L'informatica e l'intelligenza artificiale possono contribuire anche ad altre attività utili per affrontare emergenze quali quella attuale. Ad esempio:
  - (a) Chatbot e metodi basati su NLP possono essere usati per il supporto ai call center che devono gestire le emergenze;
  - (b) Sistemi per tracciare la mobilità delle persone possono aiutare a limitare la diffusione del virus;

- (c) Sistemi automatici di rilevazione della fake news possono essere d'aiuto per una corretta gestione dell'informazione

Cambiando argomento, sotto altri temi sulla Explainable AI,

5. Approfondimento sul tema della Explainable AI, illustrando le principali tecniche che si utilizzano oggi in questo ambito.
6. Definizione di un sistema che possa servire da supporto per la selezione di candidati per un corso di formazione e che in seguito possa anche servire per suggerire percorsi di apprendimento personalizzati per il singolo partecipante. Tale sistema deve anche poter rispondere a eventuali ricorsi di candidati scartati o a delucidazioni sui metodi di apprendimento suggeriti. Infine, un ultimo tema:
7. Approfondimento delle Graph Neural Network (GNN).

## 2 Progetti

I temi di cui sotto riguardano lo sviluppo di progetti reali, nei quali oltre alla descrizione dell'idea di deve anche lavorare a un parte della sua implementazione

**Big Code 1.** Sviluppo di un modello predittivo che riconosca il linguaggio il codice contenuto in un file, prendendo in input solo il contenuto del file.

**Big Code 2.** Si consideri il dataset Rosetta Code e si scelgano un certo numero di task (non inferiore a 10). Costruire un modello automatico che sia in grado di classificare i vari esempi di codice rispetto ai task che ognuno di questi dovrebbe svolgere. Si provino diverse rappresentazioni del codice (es. sequenza di caratteri, sequenza di parole, immagine del codice in un certo editor, ecc.) e diversi algoritmi di classificazione.

**Big Code 3.** ...

**Licenze.** Sviluppo di un modello predittivo che a partire dal contenuto testuale di un file riconosca la licenza software contenuta in esso

**NLP.** Si consideri il dataset Yelp Open Dataset e si utilizzi il suo contenuto per diversi obiettivi, quali ad esempio

- Riconoscimento automatico di una review positiva o negativa;
- Raggruppamento degli utenti in base alle loro preferenze o comportamento sulla piattaforma;
- Raggruppamento automatico dei locali in base a criteri di similitudine data una certa località .

Ogni altra idea di utilizzo o analisi del dataset può essere presa in considerazione.

**SDN.** Applicazione del constraint programming al problema del posizionamento delle funzioni di rete virtualizzate in ambito SDN.

**GNN** Applicazione delle Graph Neural Network (GNN) per valutare il comportamento di reti con topologie e configurazioni differenti. Si potrebbe collegare a questa challenge (<https://bnn.upc.edu/challenge2020>).

**GNN** Applicazione delle Graph Neural Network (GNN) in ambiti dove esistono grafi molto grandi, come ad esempio quello usato per l'indicizzazione dell'archivio Software Heritage.

**Scheduling.** Creare un sistema automatico di scheduling per organizzare le presentazioni di una conferenza. In particolare, ogni partecipante può esprimere un certo grado di interesse (da 0 a 5) per ogni presentazione, si hanno a disposizione  $X$  aule, ogni presentazione  $(i)$  ha la durata di  $T_i$  minuti. Si dispongano  $N$  presentazioni nelle  $Y$  giornate della conferenza dalle ore 9:00 alle 17:00, sapendo che dalle 13 alle 14 si ha la pausa pranzo, massimizzando l'interesse dei partecipanti. Si consideri che i presentatori possono avere anche più di una presentazione ciascuno.