***SocialBook***

***Progetto Fondamenti di Intelligenza Artificiale***

***(2020/2021)***

Studenti:

Barbato Alessia (0512105858)

Proietto Angelica (0512105762)

Russo Luca (0512105840)

******

# 1.Introduzione

SocialBook si propone come un social innovativo che permette la nascita di una vasta community unita dalla passione comune per i libri, dando agli utenti la possibilità di interagire tra loro.

Ogni utente, quando si registra, inserirà alcune informazioni personali di vario tipo, in particolare sui suoi interessi e preferenze letterarie, in modo da ricevere dei suggerimenti di altri utenti “simili”, ovvero utenti che appartengono allo stesso gruppo (cluster).

La suddivisione in cluster avviene grazie ad un’attenta analisi dei dati e a un lavoro di pre-processing che ci consente di individuare le caratteristiche principali su cui andare a studiare la similarità degli utenti, in modo da rendere la profilazione il più efficiente possibile.

Progettare e costruire un agente intelligente che riesca a effettuare suggerimenti basandosi su informazioni appartenenti agli ambiti più svariati (si passa infatti dagli hobby, ai generi di libri preferiti e così via) non è di certo un lavoro semplice.

Il primo punto però, consiste sicuramente nella definizione dell’ambiente, che si trova immediatamente al punto successivo.

# 2.Definizione dell’ambiente

L’ambiente viene definito tramite la rappresentazione schematica PEAS(*Performance, Environment, Actuator, Sensors*) :

• Prestazioni: Le prestazioni dell’agente vengono valutate in base all’accuratezza con cui consiglia ad un utente altri utenti a cui potrebbe essere interessato.

• Ambiente: L’ambiente in cui l’agente opera è composto da utenti registrati a una piattaforma riguardante la tematica “libri”.

• Attuatori: **//da aggiungere**

• Sensori: **//da aggiungere**

# 3.Raccolta dati

Il dataset è la collezione dei dati (campioni), ognuno con le proprie caratteristiche, che rappresentano il dominio di interesse; viene utilizzato per addestrare e successivamente testare l’agente intelligente progettato.

Per la natura dell’agente, il cui obiettivo è quello di “apprendere” e saper suddividere gli utenti in base alla similarità tra essi, si è reso necessario effettuare una sorta di indagine per poter raccogliere i dati di partenza, in modo da garantire il funzionamento efficace del modulo che abbiamo costruito.

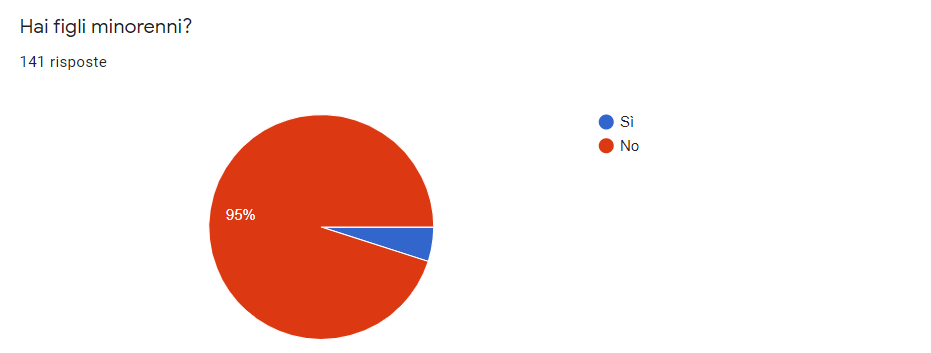
Per fare ciò, ci siamo serviti di un questionario sottoposto a quante più persone possibili, in modo da avere un dataset abbastanza ampio e non rischiare di mandare l’agente in overfitting (fenomeno che si verifica nel momento in cui un modello statistico si adatta così bene ai dati osservati da sembrare, apparentemente, perfetto sui dati di training, salvo che, quando si va ad applicare il modello sui dati di test, si registrano molti errori).

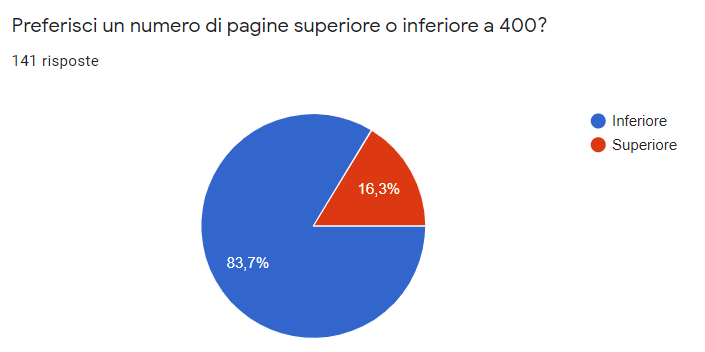
Il nostro questionario ([Socialbook (google.com)](https://docs.google.com/forms/d/1m8v-GKTujO15kvwWOejnmVg1X4RbXxKW61YECixXLTg/viewform?edit_requested=true)) raccoglie informazioni sia personali che per quanto riguarda l’ambiente letterario.

Il lettore, infatti, è caratterizzato dalle seguenti informazioni:

Nella fase di stesura del questionario, infatti, il team si è soffermato su varie informazioni, spaziando dalla fascia d’età, agli hobby, ai generi letterari prediletti, alla quantità di libri letti all’anno etc.

Purtroppo, non tutti i dati raccolti sono stati significativi e utili per il problema in questione, a causa di una scarsa distribuzione delle risposte. Di seguito, sono stati allegati degli esempi.





Per questo motivo, si è deciso di filtrare il dataset di origine, con lo scopo di ridurre lo spazio di dimensione dei dati in input, andando a considerare solo le caratteristiche realmente utili per la risoluzione del problema.

AGGIUNGERE GRAFICO ETA E OCCUPAZIONE

Poiché per scelta libro la risposta preponderante è per genere, allora consideriamo GENERE e così selezioniamo anche hobby