



**Università degli Studi di Salerno**  
Documentazione progetto Fondamenti di  
Intelligenza Artificiale

<i>Autore</i>	<i>Matricola</i>
Costante Luigina	0512110457
Lo Conte Simona	0512110922
Napolillo Marta	0512109836

# Indice

<b>1</b>	<b>Introduzione</b>	<b>2</b>
<b>2</b>	<b>Business Understanding</b>	<b>3</b>
2.1	Obiettivi di business . . . . .	3
2.2	PEAS . . . . .	3
2.3	Proprietà dell'ambiente . . . . .	3
2.4	Analisi del problema . . . . .	4
<b>3</b>	<b>Data Understanding</b>	<b>5</b>
3.1	Acquisizione dei dataset . . . . .	5
3.2	Analisi dei dataset . . . . .	5
<b>4</b>	<b>Data Preparation</b>	<b>6</b>
4.1	Data Cleaning . . . . .	6
4.2	Feature Scaling . . . . .	6
4.3	Feature Selection . . . . .	6
4.4	Data Balancing . . . . .	6
<b>5</b>	<b>Modeling</b>	<b>7</b>
5.1	Scelta dell'algoritmo da utilizzare . . . . .	7
5.2	Fase di addestramento . . . . .	7
<b>6</b>	<b>Evaluation</b>	<b>8</b>
6.1	Elbow point . . . . .	8
6.2	Silhouette coefficient . . . . .	8
6.3	MoJo distance . . . . .	8
<b>7</b>	<b>Deployment</b>	<b>9</b>
<b>8</b>	<b>Glossario</b>	<b>10</b>

# 1 Introduzione

## 2 Business Understanding

### 2.1 Obiettivi di business

### 2.2 PEAS

Specifica PEAS dell'ambiente.

<b>Performance</b>	Capacità dell'agente di suggerire all'utente contenuti che rispecchiano i suoi gusti.
<b>Environment</b>	L'ambiente in cui l'agente opera è rappresentato da iLike, un'applicazione in cui gli utenti possono scrivere recensioni ed esprimere preferenze sui contenuti che si trovano all'interno di essa.
<b>Actuators</b>	Risposta del Conversational Agent.
<b>Sensors</b>	Utterances (messaggi in linguaggio naturale dati in input al CA da un utente umano).

### 2.3 Proprietà dell'ambiente

L'ambiente possiede le seguenti proprietà:

- **Completamente osservabile:** l'agente ha accesso all'elenco dei contenuti presenti nell'applicazione e alle preferenze degli utenti in qualsiasi momento;
- **Stocastico:** lo stato dell'ambiente varia indipendentemente dall'azione intrapresa dall'agente;
- **Sequenziale:** le decisioni prese dall'agente dipendono dalle azioni passate dell'utente;
- **Statico:** nel momento in cui l'agente sta elaborando la sua decisione l'utente non può modificare le sue preferenze;
- **Discreto:** i suggerimenti dati dall'agente dipendono dalla combinazione di contenuti preferiti di cui l'utente dispone ed esistono un numero limitato di possibili combinazioni;
- **Singolo-agente:** esiste un unico agente che opera nell'ambiente.

## 2.4 Analisi del problema

Il problema che l'agente intelligente dovrà risolvere consiste nel suggerire film/serie TV da vedere, libri da leggere ed album musicali da ascoltare in base ai contenuti presenti nei dataset dell'applicazione e soprattutto in base alle preferenze espresse dagli utenti.

Il problema in esame può essere risolto con un algoritmo di apprendimento in quanto consiste nel migliorare l'esecuzione di un task (T=fornire suggerimenti personalizzati) rispetto ad una misura di prestazione (P= numero di suggerimenti accettati dall'utente) e sulla base dell'esperienza (E= database di contenuti non etichettati). Inoltre l'algoritmo di apprendimento in questione è di tipo non supervisionato in quanto non si dispone di un database contenente dati già etichettati, bensì dovrà essere l'agente capace di apprendere il valore reale della variabile dipendente sulla base delle conoscenze di cui dispone.

Nello specifico il problema in esame può essere risolto tramite l'utilizzo di un algoritmo di clustering. Una volta che l'utente ha espresso le sue preferenze riguardanti contenuti presenti nell'applicazione, l'algoritmo creerà, in base ad una misura di similarità (che verrà definita in seguito), dei cluster contenenti film, serie TV, libri e album musicali dotati di un certo grado di omogeneità. Procederà quindi a consigliare nuovi contenuti in base alla clusterizzazione effettuata.

I suggerimenti verranno dati solo qualora l'utente ne faccia richiesta ed il tutto avviene in maniera automatica tramite l'utilizzo di un Conversational Agent.

## 3 Data Understanding

### 3.1 Acquisizione dei dataset

Durante la scelta dai dati da fornire al machine learning le possibili scelte da seguire erano sostanzialmente due:

- Creare un dataset contenente gli utenti di iLike ed analizzare il loro comportamento, al fine di creare cluster di utenti i quali hanno preferenze simili;
- Cercare dataset con le informazioni sui contenuti e creare cluster di contenuti.

I problemi riscontrati sono:

- La disponibilità di dati era maggiore nei dataset già esistenti;
- Ogni utente ha gusti differenti, quindi la similarità dei contenuti in un cluster può non essere sempre veritiera;
- Individuare dataset con un numero ottimale di istanze e le giuste informazioni sui contenuti richiede un'accurata analisi.

Al seguito di una breve trade-off abbiamo preferito di utilizzare dataset già creati poichè, la disponibilità di dati e la giusta similarità di elementi in un cluster, agevola le prestazioni dell'algoritmo di machine learning.

### 3.2 Analisi dei dataset

I dataset reperiti sulla piattaforma Kaggle sono:

- [Film](#)
- [Serie TV](#)
- [Album Musicali](#)
- [Libri](#)

## 4 Data Preparation

### 4.1 Data Cleaning

### 4.2 Feature Scaling

### 4.3 Feature Selection

### 4.4 Data Balancing

## 5 Modeling

5.1 Scelta dell'algoritmo da utilizzare

5.2 Fase di addestramento



## 6 Evaluation

6.1 Elbow point

6.2 Silhouette coefficient

6.3 MoJo distance

## 7 Deployment

## 8 Glossario

<b>Contenuto</b>	Elemento appartenente all'insieme di film, serie TV, libri e album musicali.
<b>Cluster</b>	Sottoinsieme di contenuti con caratteristiche simili.
<b>Machine Learning</b>	È la branca dell'Intelligenza Artificiale che include tutti gli algoritmi che possano imparare dai dati e sulla base di questi fare previsioni.