



Modulo di FIA

Fondamenti di Intelligenza Artificiale

BiblioNet

Riferimento	C07_Modulo_FIA_ver.1
Versione	1.0
Data	06/01/2021
Destinatario	Prof. Fabio Narducci Prof. Fabio Palomba
Presentato da	Nicola Pagliara Viviana Pentangelo Gianmario Voria
Approvato da	Stefano Lambiase



Team members

Nome	Ruolo nel progetto	Acronimo	Informazioni di contatto
Stefano Lambiase	Project Manager	SL	s.lambiase7@studenti.unisa.it
Gianmario Voria	Team Member	GV	g.voria6@studenti.unisa.it
Viviana Pentangelo	Team Member	VP	v.pentangelo4@studenti.unisa.it
Nicola Pagliara	Team Member	NP	n.pagliara1@studenti.unisa.it



Revision History

Data	Versione	Descrizione	Autori
06/01/2021	0.1	Prima stesura	GV, NP, VP
06/01/2021	0.2	Stesura dei paragrafi 1.1, 1.2, 2.1, 2.2, 3.2	GV, NP, VP
02/03/2021	0.3	Revisione paragrafo 1.1, 1.2	GV, VP
04/03/2021	0.4	Stesura dei paragrafi 1.3, 3.1, 3.2	VP
05/03/2021	0.5	Stesura paragrafo 4.1	NP
08/03/2021	0.6	Stesura paragrafo 2.3	GV
10/03/2021	0.7	Stesura dei paragrafi 4.2, 4.3, 4.4	GV, VP
10/03/2021	0.8	Stesura paragrafi 5.1, 6	NP
18/03/2021	1.0	Revisione	GV, NP, VP



Sommario

Team members	2
Revision History	3
1 Introduzione	6
1.1 Obiettivo del sistema.....	6
1.2 Obiettivo dell'agente intelligente	7
1.3 Organizzazione del documento	7
2 Analisi del problema	8
2.1 Formalizzazione del problema	8
2.2 PEAS	8
2.2.1 Performance	8
2.2.2 Environment.....	8
2.2.3 Actuators	9
2.2.4 Sensors.....	9
2.3 Proprietà dell'ambiente.....	9
3 Dataset	10
3.1 Raccolta dei dati.....	10
3.2 Struttura del Dataset.....	12
4 Soluzione proposta.....	13
4.1 Modello	13
4.1.1 Bayesian Model	13
4.1.2 Applicazione del Modello nel caso BiblioNet	14
4.1.3 Motivazioni sulla scelta del modello	15
4.2 Tecnologie ed implementazione del modello	16
4.3 Integrazione in BiblioNet	16
4.4 Risultati e conclusioni.....	18



Laurea Triennale in informatica-Università di Salerno
Corso di *Fondamenti di Intelligenza Artificiale*
Prof F. Narducci, Prof F. Palomba

5	Sviluppi futuri	18
6	Glossario	19



1 Introduzione

Il presente documento consiste nella documentazione del modulo di Intelligenza Artificiale della piattaforma BiblioNet.

Nella sezione di introduzione verranno descritti il sistema e i suoi obiettivi, conseguentemente gli obiettivi dell'agente intelligente, e una panoramica sul documento.

Al seguente link è disponibile l'intera documentazione del sistema, che contiene anche l'evoluzione del modulo in fase di Requirements Analysis e System Design:

[Documenti Biblionet](#)

Il modulo è stato implementato su un fork del progetto originale, la cui repository è disponibile al seguente link:

[Biblionet Modulo FIA](#)

In particolare, all'interno della cartella, l'implementazione effettiva può essere raggiunta tramite il seguente link:

[/biblionet/tree/AI-Module/src/main/java/it/unisa/c07/biblionet/moduloIntelligenzaArtificiale](#)

1.1 Obiettivo del sistema

Il sistema che si vuole realizzare ha come obiettivo principale quello di facilitare l'affaccio a nuove persone al mondo della lettura. Attraverso una piattaforma online, viene fornito al lettore la possibilità di cercare le biblioteche a lui più vicine e di visionare in modo semplice e veloce la lista dei libri disponibili in esse.

Il sistema ha un duplice compito, ottimizzare il lavoro delle biblioteche che si interfacciano con i loro clienti lettori e, allo stesso tempo, aiutare i lettori meno esperti ad approcciarsi a questo mondo.

In particolare, esso permette di:

- Conoscere le biblioteche della zona in modo facile e veloce
- Prenotare un libro a distanza
- Fornire un canale di contatto rapido con le biblioteche

Inoltre, il sistema offre la possibilità di organizzare eventi, anche in collaborazione con le scuole, con l'obiettivo di invogliare la popolazione a partecipare ad attività di lettura.



Nel dettaglio, la figura dell'esperto che lavora in una biblioteca è molto importante, e si punta a crearne sempre di più. Infatti, grazie ad essi i lettori potranno accrescere le loro passioni e rendere più viva la loro esperienza di lettura con la partecipazione ad attività di gruppo collettive organizzate.

Il sistema nel complesso, tramite le funzionalità precedentemente citate, punta a facilitare il lavoro svolto dai bibliotecari, aumentando allo stesso tempo la loro produttività, e ad incentivare la popolazione ad avvicinarsi alla lettura o ad approfondirla.

1.2 Obiettivo dell'agente intelligente

Nell'ambito della piattaforma Biblionet, viene dato per scontato che un generico utente si voglia registrare come Lettore, per poter usufruire dei servizi offerti. In quanto lettore, dovrebbe essere già a conoscenza delle proprie preferenze di lettura, che rappresentano un'informazione molto importante all'interno del sistema.

Ciò non lascia molto spazio all'accoglienza di nuovi utenti che vogliono avvicinarsi al mondo della lettura, ma che non sono certi dei propri interessi o delle proprie preferenze. Questi ultimi potrebbero trovarsi in difficoltà, ad esempio, nell'iscrizione ad un club del libro, che si basa su libri accomunati dagli stessi generi, o a compilare le proprie preferenze di lettura.

Per risolvere questo problema, viene proposto ai nuovi iscritti un Questionario di Supporto: si tratta di un modulo di Intelligenza Artificiale che, sulla base delle risposte date dall'utente ad un breve questionario di carattere generale, propone una serie di generi letterari dai quali potrebbe cominciare.

L'obiettivo dell'agente è dunque quello di raccogliere ed elaborare informazioni sull'utente, sulla base delle risposte ricevute, per essere in grado di fare una predizione su quali potrebbero essere i generi letterari di suo interesse.

L'agente intelligente sarà quindi un agente capace di apprendere e, sulla base delle informazioni ottenute, riuscirà a produrre una soluzione al problema proposto.

1.3 Organizzazione del documento

Nelle sezioni successive si procederà con l'analisi del problema individuato, procedendo con la sua formalizzazione e descrizione dell'ambiente in cui opera.

Una sezione sarà dedicata alle modalità scelte per il raccoglimento dei dati e alla creazione e struttura del Dataset risultante.



In seguito, si descriverà la soluzione proposta, analizzando modelli, tecnologie e architetture utilizzati per tale scopo.

2 Analisi del problema

In questa sezione viene analizzato il problema individuato, la sua formalizzazione e la sua descrizione mediante il modello PEAS. Oltre ciò, viene descritto l'ambiente in cui l'agente opera, e le sue principali caratteristiche.

2.1 Formalizzazione del problema

- **Stato iniziale:** Una serie di risposte date dall'utente e dei generi letterari.
- **Azioni:** L'agente può selezionare una risposta dell'utente associandola ad un genere, mediante un dataset per applicare un modello alle suddette risposte.
- **Modello di transizione:** Restituisce uno o più generi associati alla risposta dell'utente ad una domanda.
- **Insieme degli stati:** L'insieme dell'associazione delle risposte alle domande del questionario e la lista dei generi risultanti.
- **Test obiettivo/Funzioni obiettivo:** Verificare se il genere è conforme alle risposte.

2.2 PEAS

2.2.1 Performance

La misura di performance del modulo si basa sulla accuratezza di associazione tra le risposte date da un lettore ed il genere prodotto come risultato. Questa associazione viene fatta attraverso una classificazione delle risposte in caratteristiche descrittive dei generi candidati, dalle quali verranno poi selezionati i generi mostrati all'utente.

2.2.2 Environment

In quanto facente parte del sistema Biblionet, ovvero una web application, l'ambiente del modulo consiste in tutte le funzionalità offerte dalla piattaforma ed in particolare gli elementi che ne permettono l'utilizzo sul browser. Fanno parte dell'ambiente tutti gli elementi di visualizzazione che permettono l'interazione con l'utente, ai quali si



aggiungono le informazioni relative al questionario che l'utente stesso compila e i possibili generi letterari.

2.2.3 Actuators

Gli attuatori consistono essenzialmente in tutto ciò che fa parte della logica di business della web application, ossia le funzioni che permettono il calcolo del risultato e la corretta visualizzazione di quest'ultimo.

2.2.4 Sensors

I sensori sono rappresentati da tutti i boundary objects che permettono l'interazione dell'utente col sistema, in particolare i bottoni per richiedere la compilazione del questionario di supporto e gli elementi visuali utilizzati per rispondere alle domande.

2.3 Proprietà dell'ambiente

- Completamente osservabile: i sensori dell'agente gli permettono di ricevere sempre lo stato completo dell'ambiente, in quanto verrà attivato solo dopo aver risposto a tutte le domande del questionario.
- Stocastico: dato che la risposta dell'agente è un semplice consiglio, lo stato dell'ambiente, ovvero come l'utente decide di interagire con Biblionet, dipende soprattutto dalle decisioni di quest'ultimo. Per tale motivazione, lo stato dell'ambiente non è completamente determinato dalle azioni dell'agente.
- Episodico: l'agente delibera un risultato solo quando viene richiesto dall'utente in maniera episodica.
- Statico: durante il calcolo della risposta, il controller della web application mette l'utente in attesa del risultato, quindi non è possibile modificare lo stato dell'ambiente.
- Discreto: l'insieme delle possibili percezioni sono le domande e le risposte stabilite in maniera preliminare, quindi a tutti gli effetti un numero limitato.
- Agente Singolo: nell'ambiente è presente un solo agente intelligente.



3 Dataset

In questa sezione viene spiegato il processo effettuato per la costruzione del Dataset utilizzato dall'agente intelligente.

Verrà spiegata la modalità di raccolta dei dati e la loro formattazione nel Dataset risultante.

3.1 Raccolta dei dati

Data la natura del problema ed il tipo di apprendimento utilizzato dall'agente per effettuare le sue predizioni, per la raccolta dei dati si è scelto di collezionare le risposte alle domande del questionario da una serie di utenti già consci delle proprie preferenze in ambito letterario.

Si è quindi deciso di costruire un questionario da sottoporre agli utenti, strutturato in due sezioni:

1. Una serie di 25 domande a risposta multipla, con 4 scelte possibili per ciascuna domanda. Le domande sono state strutturate per coprire vari ambiti di carattere generale, spaziando fra gusti musicali, interessi, abitudini ed un'approssimativa profilazione psicologica.

<i>Quali sono le tue attività preferite da svolgere nel tempo libero?</i>
<i>Qual è il tuo genere musicale preferito?</i>
<i>Quale tipo di materie ti interesserebbe più studiare, studi o hai studiato?</i>
<i>Cosa ti infastidisce di più?</i>
<i>Da che tipo di storie ti lasci coinvolgere di più?</i>
<i>Quanto ti appassionano i misteri e gli argomenti d'inchiesta?</i>
<i>Quali di queste è una sensazione che ti piace provare?</i>
...

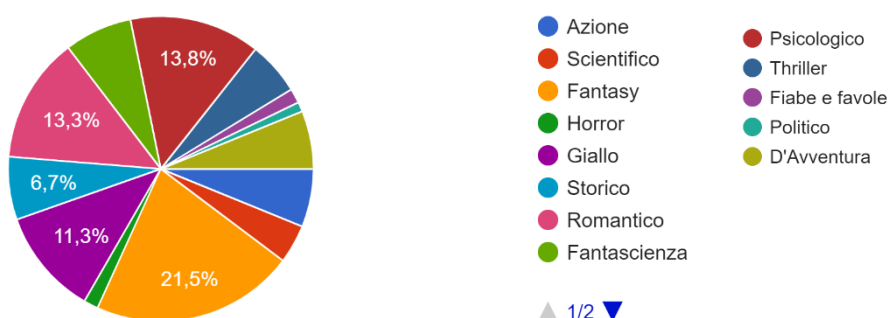
Esempi di domande presenti nel questionario

2. La scelta di un singolo genere preferito da una lista di quelli presenti su Biblionet, e successivamente una serie di tre aggettivi per descriverlo fra quelli proposti.

Di seguito vengono mostrati i risultati della seconda sezione del questionario, sulla base delle 195 risposte raccolte

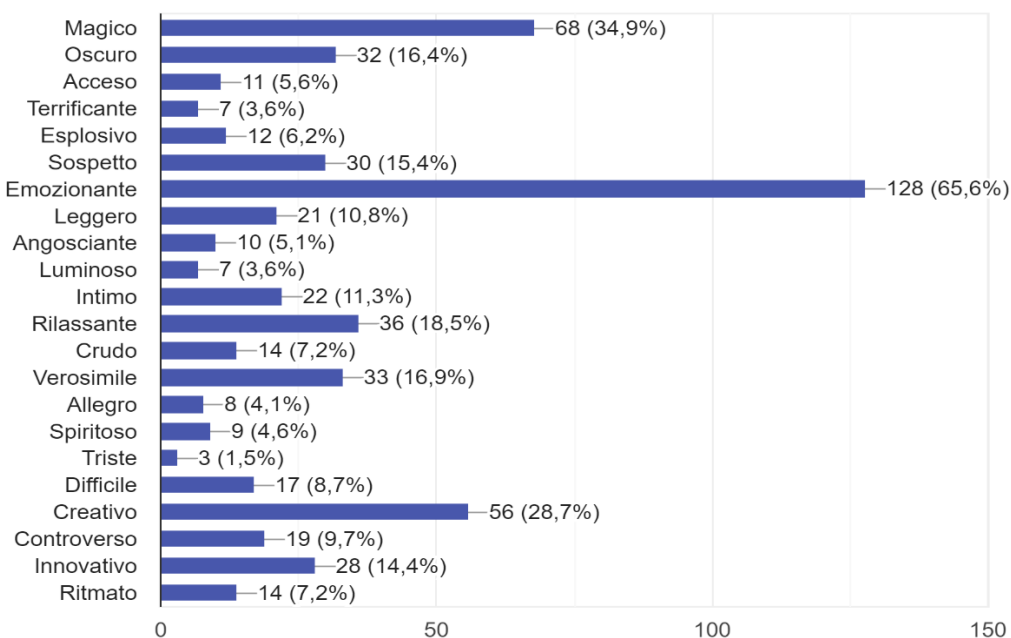
Qual è il tuo genere letterario preferito?

195 risposte



Scegli 3 aggettivi che meglio descrivono, secondo te, il genere che hai scelto

195 risposte



3.2 Struttura del Dataset

Successivamente alla raccolta dei dati, si è proceduto alla loro formattazione in un documento Excel da cui l'agente intelligente potesse allenarsi e apprendere.

Lo strumento utilizzato per la somministrazione del questionario, Google Moduli, ha fornito in maniera automatica un documento di Google Fogli così strutturato:

	Risposta1	Risposta2	...	Risposta25	Genere selezionato	Lista di caratteristiche
Utente1						
...						
Utente 195						

Nella tabella sopra riportata ogni riga rappresenta una risposta completa al questionario da parte di un utente.

Da questo, sommariamente già correttamente formattato, per ottenere la forma desiderata dei dati in nostro possesso, è stato costruito il documento Excel finale rappresentante il dataset, così strutturato:

	Risposta 1	Risposta 2	...	Risposta 25	Caratteristica 1	Caratteristica 2	Caratteristica 3	Genere
Utente 1								
...								
Utente 195								

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	/	Sognatore e intensi e caldi	Colori intensi e caldi	Spaventarsi per un jumpscar e ben	Grifondoro	Black, Br	Coccolato a caldo, tè o cappuccino	Non li piacciono molto, ma li piace il country	Indie pop, indie rock, o country	Luoghi ricchi di cultura e storia, medicina e	Materie basate sulle matematiche, la fisica, la	COU, Counter Strike, Apex Legends, Skyrim, Final Fantasy, Dark Souls, The Legend of Zelda	Foresta d'autunno	Nessuno	La confusione e i rumori	SI, forse anche troppo	Una cena o un pranzo fuori	Urra e molto difficile che mi interessi a qualcosa	Verbrevisim o tempo o quasi mai	Con un po' di zucchero	Crede che esistono fenomeni	Crede che rigornali e	Storie che aggiungono ad un contesto	Le mie scelte sono maggiormente istintive, ma	Terra	Oscura	Angosciante/erosomili	Scientifico	
2	programmi	Anticurioso, curioso, pianificato	Colori freddi e scuri	Arrivare alla scoperta di nuove	Serpeverde	error, Dark, T	Un caffè	Non li trovo interessanti, ma li trovo utili	Dance, house o techno	Una o due settimane in un paese	Materie basate sulle matematiche, la fisica, la	Dark Souls, The Legend of Zelda	Città di notte	Amico	L'agire con qualcuno a cui la monotonia	No, per nulla	Tecnologia	Urra e molto difficile che mi interessi a qualcosa	Abbassati, ma se intervallati da	Amaro	Assolutamente no	Crede che rigornali e	Storie che aggiungono ad un contesto	Le mie scelte sono maggiormente istintive, ma	Aria	Magico	Emozionante	Creativo	Fantasy
3	programmi	Labovvero o, socievole ed	Colori vivaci e saturi	Forse adrenalinica	Serpeverde	ow I Met Yok	Drink alcolico o una birra	Non li trovo interessanti, ma li trovo utili	Metal, rock, o rap	Una o due settimane in un paese	Materie basate sulle matematiche, la fisica, la	Prato	Amico	La monotonia e la ripetitività	No, per nulla	Tecnologia	Urra e molto difficile che mi interessi a qualcosa	Abbassati, ma se intervallati da	Non bevo caffè	Assolutamente no	Crede che rigornali e	Storie che aggiungono ad un contesto	Le mie scelte sono maggiormente istintive, ma	Fuoco	Magico	Terrificante	Creativo	Fantasy	
4	Arte, musica, crafting	Sognatore e, creativo, empatico	Colori intensi e caldi	Forse adrenalinica	Grifondoro	W, Gine	Coccolato a caldo, tè o cappuccino	Non li piacciono molto, ma li piace il country	Indie pop, indie rock, o country	L'estate in una località di mare, a Venezia	Materie basate sulle matematiche, la fisica, la	The Sims, Darkwing G, Animal Crossing	Foresta d'autunno	Amico	La monotonia e la ripetitività	SI, forse anche troppo	Una cena o un pranzo fuori	Urra e molto difficile che mi interessi a qualcosa	Abbassati, ma se intervallati da	Con un po' di zucchero	SI, credo molto nei paradossi e alle	MI ritengo una persona ambiziosa e non mi arrendo mai	Storie con avventure e ambientazioni alla stregua di quelle che si raccontano in un modo molto crudo	Le mie scelte sono maggiormente istintive, ma	Aria	Magico	Oscur	Accesso	Azione
5	Arte, musica, crafting	Sognatore e, creativo, empatico	Colori intensi e caldi	Ridere a crepapelle	Tassoross	error, Dark, T	Coccolato a caldo, tè o cappuccino	Non li trovo interessanti, ma li trovo utili	Indie pop, indie rock, o country	Luoghi ricchi di cultura e storia, medicina e	Materie basate sulle matematiche, la fisica, la	COU, Counter Strike, Apex Legends, Skyrim, Final Fantasy, Dark Souls, The Legend of Zelda	Foresta d'autunno	Amico	La confusione e i rumori	SI, forse anche troppo	Una cena o un pranzo fuori	Urra e molto difficile che mi interessi a qualcosa	Verbrevisim o tempo o quasi mai	Con un po' di zucchero	Crede che esistono fenomeni	Crede che rigornali e	Storie che aggiungono ad un contesto	Le mie scelte sono maggiormente istintive, ma	Terra	Oscura	Angosciante/erosomili	Scientifico	
6	Arte, musica, crafting	Sognatore e, creativo, empatico	Colori intensi e caldi	Spaventarsi per un jumpscar e ben	Grifondoro	Black, Br	Coccolato a caldo, tè o cappuccino	Non li trovo interessanti, ma li trovo utili	Indie pop, indie rock, o country	Luoghi ricchi di cultura e storia, medicina e	Materie basate sulle matematiche, la fisica, la	COU, Counter Strike, Apex Legends, Skyrim, Final Fantasy, Dark Souls, The Legend of Zelda	Foresta d'autunno	Amico	La confusione e i rumori	SI, forse anche troppo	Una cena o un pranzo fuori	Urra e molto difficile che mi interessi a qualcosa	Verbrevisim o tempo o quasi mai	Con un po' di zucchero	Crede che esistono fenomeni	Crede che rigornali e	Storie che aggiungono ad un contesto	Le mie scelte sono maggiormente istintive, ma	Terra	Oscura	Angosciante/erosomili	Scientifico	
7	programmi	Sognatore e, creativo, empatico	Colori intensi e caldi	Arrivare alla scoperta di nuove	Tassoross	ow Black, Br	Un caffè	Non li trovo interessanti, ma li trovo utili	Dance, house o techno	Una o due settimane in un paese	Materie basate sulle matematiche, la fisica, la	Dark Souls, The Legend of Zelda	Città di notte	Amico	L'agire con qualcuno a cui la monotonia	No, per nulla	Tecnologia	Urra e molto difficile che mi interessi a qualcosa	Abbassati, ma se intervallati da	Amaro	Assolutamente no	Crede che rigornali e	Storie che aggiungono ad un contesto	Le mie scelte sono maggiormente istintive, ma	Aria	Magico	Emozionante	Creativo	Fantasy
8	Arte, musica, crafting	Sognatore e, creativo, empatico	Colori intensi e caldi	Spaventarsi per un jumpscar e ben	Grifondoro	Black, Br	Coccolato a caldo, tè o cappuccino	Non li trovo interessanti, ma li trovo utili	Indie pop, indie rock, o country	Luoghi ricchi di cultura e storia, medicina e	Materie basate sulle matematiche, la fisica, la	COU, Counter Strike, Apex Legends, Skyrim, Final Fantasy, Dark Souls, The Legend of Zelda	Foresta d'autunno	Amico	La confusione e i rumori	SI, forse anche troppo	Una cena o un pranzo fuori	Urra e molto difficile che mi interessi a qualcosa	Verbrevisim o tempo o quasi mai	Con un po' di zucchero	Crede che esistono fenomeni	Crede che rigornali e	Storie che aggiungono ad un contesto	Le mie scelte sono maggiormente istintive, ma	Terra	Oscura	Angosciante/erosomili	Scientifico	

Righe di esempio del dataset

4 Soluzione proposta

Per la soluzione si è scelto un apprendimento di tipo statistico. Il motivo di questa scelta risiede nel fatto che il problema è composto da poche variabili, discrete, e queste ultime rappresentano correttamente l'ambiente e il tipo di agente che si voleva ottenere. Inoltre, data la costruzione di un dataset molto specifico e basato sulle esperienze di utenti vicini al target della piattaforma, si è pensato che una soluzione basata sull'inferenza probabilistica sarebbe stata adeguata, considerando anche la semplicità dell'utilizzo di questi modelli rispetto ad altri tipi di apprendimento.

4.1 Modello

Il modello scelto per la soluzione proposta è la **Rete Bayesian**a. Nel presente paragrafo verrà descritto il modello e verrà illustrato come le proprietà e le caratteristiche di quest'ultimo sono state rispettate. Inoltre, saranno discusse le motivazioni che hanno portato alla scelta di questo modello.

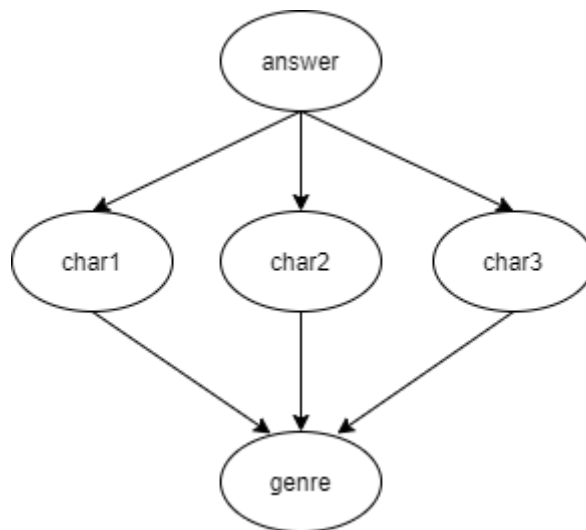
4.1.1 Bayesian Model

Il modello Bayesian si basa sul concetto di Rete Bayesian, definita come un grafo orientato in cui ogni nodo è etichettato con un'informazione probabilistica quantitativa.

La sua specifica prevede che:

1. I nodi della rete siano costituiti da un insieme di variabili casuali, che possono essere discrete o continue. Nel nostro caso, le variabili casuali sono:
 - risposta dell'utente, che chiameremo **answer**,
 - tre caratteristiche che possano descrivere il genere: **char1, char2, char3**,
 - genere letterario, **genre**.
2. Un insieme di archi orientati collega coppie di nodi. Se c'è un arco da X a Y, X è genitore di Y. La soluzione prevede tre archi che collegano **answer** a **char1, char2 e char3**, e tre che collegano **char1, char2 e char3** a **genre**.
3. Ogni nodo X_i ha una distribuzione di probabilità condizionata del tipo $P(X_i | \text{Genitori}(X_i))$. Sulla base della rete costruita, ad esempio, le CPT di **char1** saranno condizionate dalla variabile **answer**, e le CPT di **genre** saranno condizionate dalle variabili **char1, char2 e char3**.
4. Il grafo è un DAG, cioè un grafo orientato con assenza di cicli. Nel paragrafo 4.1.2 è mostrata la rappresentazione grafica della rete, che dimostra la precedente affermazione.

4.1.2 Applicazione del Modello nel caso BiblioNet



La figura mostra la topologia della rete utilizzata per l'ambiente BiblioNet.

Prima di costruire la rete, abbiamo eseguito l'identificazione delle variabili causali e la modellazione delle dipendenze di quest'ultime, basandosi sui risultati ottenuti grazie alla formalizzazione ed al PEAS.

La scelta delle variabili causali *answer* e *genre* è legata direttamente ai risultati dell'analisi del problema, dato che nella sua formalizzazione entrambe le variabili rappresentano lo **stato iniziale**.

Le variabili *char1*, *char2* e *char3* sono state introdotte per permettere al modello di eseguire l'associazione tra le risposte utente ed il genere letterario da restituire. Esso aiuta l'agente intelligente di eseguire le **azioni** sull'ambiente, sulla base di quanto definito nel **modello di transazione**. Inoltre, l'introduzione di queste variabili è stata utile per mantenere uno stato di indipendenza tra le variabili *answer* e *genre* date le caratteristiche.

Avendo a disposizione le variabili, è stato costruito manualmente il DAG rappresentante la rete, effettuando a tutti gli effetti parameter learning, in quanto il modello ha dovuto solo apprendere dai dati.

Data la tipologia di problema, la rete ha eseguito l'apprendimento grazie ad un dataset costruito in maniera artificiale.

Per effettuare l'apprendimento sui dati, quindi per il calcolo delle CPT, abbiamo scelto di utilizzare un **BayesianEstimator** piuttosto che un **MaximumLikelihoodEstimator**, in quanto il primo in presenza di un dataset piccolo e composto da poche variabili ha portato ad un overfitting minore rispetto al secondo, dato che tende a non assegnare probabilità nulla a valori poco presenti nei dati, evitando l'esclusione a priori di alcune possibilità.

4.1.3 Motivazioni sulla scelta del modello

Di seguito è riportata una tabella che esplicita i pro e i contro che sono stati valutati nella scelta del modello Bayesiano.

Vantaggi	Svantaggi
Flessibile e naturale per modellare dipendenze ed indipendenze tra le variabili causali.	Inefficiente su dataset molto complessi e di grandi dimensioni.
Chiara ed efficiente semantica probabilistica.	Più incline ad overfitting, confrontato ad altri modelli.
Efficiente su di un dataset ristretto.	

Sulla base di queste considerazioni, abbiamo ritenuto corretto l'utilizzo del modello per le seguenti motivazioni:

- Essendo il nostro problema composto da poche variabili e da dipendenze molto forti, la rete bayesiana rappresentava bene il dominio della soluzione.
- La costruzione del modello era semplice poiché avevamo un dataset molto piccolo per cui la rete si è dimostrata efficiente.
- La rete è stata costruita su di un dataset artificiale, il quale ha semplificato ancora di più la costruzione. La corretta formattazione del dataset, infatti, ha permesso un'estrazione dei vincoli facile e chiara.
- Considerato che il compito dell'agente è quello di consigliare generi, è quindi sostanzialmente privo di responsabilità. Si è deciso, quindi, di accettare la possibilità di bassa accuracy causata dall'overfitting, a favore della semplicità del modello.



4.2 Tecnologie ed implementazione del modello

Per l'implementazione è stato utilizzato il linguaggio Python, in quanto offre una grande scelta di librerie e funzionalità nell'ambito del machine learning.

Per l'implementazione del modello bayesiano è stata scelta la libreria [pgmpy](#), che mette a disposizione il suddetto modello tramite la classe BayesianModel. L'utilizzo di questa libreria ci ha permesso di definire con facilità le dipendenze tra le variabili, offrendo tutti i metodi necessari all'apprendimento del modello, la verifica della correttezza e la predizione di nuovi valori.

Per la rappresentazione del dataset è stata utilizzata la libreria [pandas](#), che ci ha permesso di leggere il file excel contenente i dati in un DataFrame leggibile dal modello.

Il modello ha quindi letto i dati sottoforma di domanda, caratteristiche e genere dal dataset e si è allenato sul 70% di essi, tramite il metodo BayesianModel.fit(), riservando il restante 30% per il testing.

Il metodo BayesianModel.checkModel() ci ha assicurato che le proprietà fondamentali della rete bayesiana fossero rispettate.

Una volta allenato, il modello viene reso persistente in modo da poter essere letto in qualsiasi momento per effettuare le predizioni.

Il modulo, quindi, prende in input cinque stringhe, che sono le risposte date da un utente al questionario, effettua la predizione grazie al metodo BayesianModel.predict() sulle risposte e restituisce i generi che ha calcolato.

4.3 Integrazione in BiblioNet

L'architettura scelta per l'integrazione del modello nella piattaforma BiblioNet si basa sull'utilizzo del modulo come **web service**.

La piattaforma Biblionet è una web application che si poggia sul framework **Spring**, scritta quindi in Java ed hostata tramite un web server su **localhost:8000**.

Essendo il modulo di intelligenza artificiale scritto in Python, una tecnologia diversa da quella della piattaforma, è stato pensato di mettere a disposizione una API invocabile tramite chiamata HTTP, che permettesse alle due tecnologie di comunicare attraverso un linguaggio comune, ovvero JSON.

Per far ciò ci siamo serviti del design pattern dell'Adapter, il cui scopo è quello di effettuare le richieste HTTP ed elaborare il JSON di risposta in un oggetto riconoscibile dalla piattaforma.

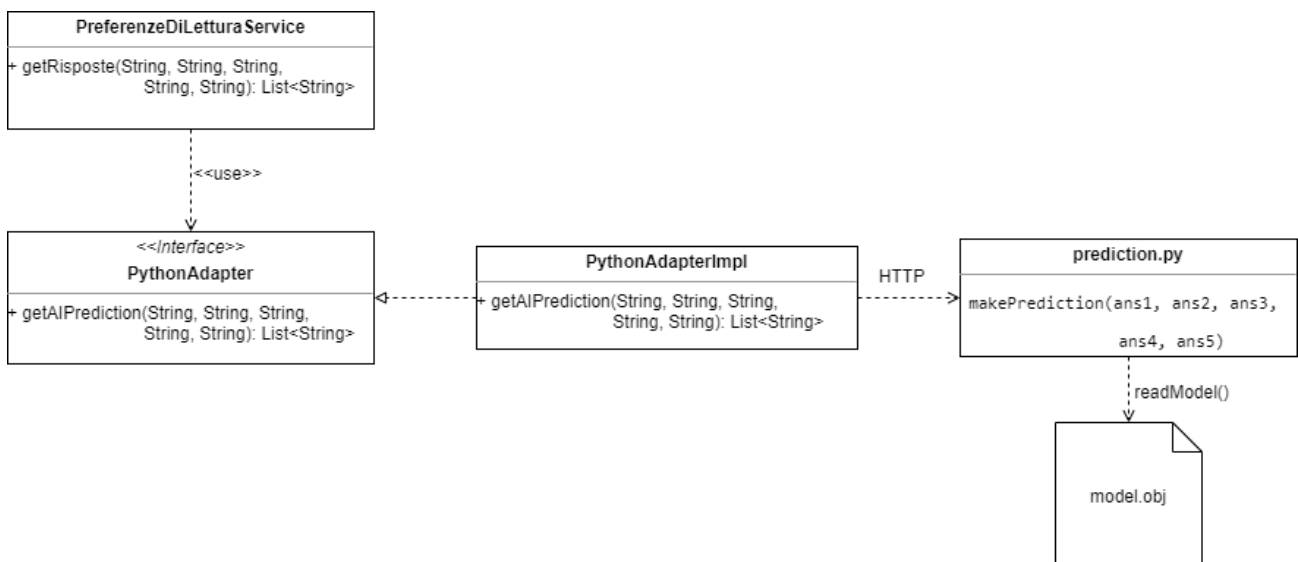
Questa scelta nasce da due motivi fondamentali:

1. è stato possibile effettuare l'integrazione del modulo senza alcun interprete fra i due linguaggi. In questo modo, Java non ha infatti bisogno di interpretare lo script in Python: le due parti riescono comunicare scambiandosi oggetti JSON, formato comprensibile ad entrambe;
2. il modulo è stato reso di fatto indipendente dalla piattaforma, slegando la sua implementazione da quest'ultima, riducendo al minimo l'accoppiamento fra i due e favorendone un eventuale riutilizzo futuro.

Per implementare quanto appena detto è stata realizzata una semplice API in Python, la quale legge il modello trainato da un file e gli fa effettuare una predizione ricevendo in input le 5 risposte. Per questo scopo ci siamo serviti di [flask](#), un micro-framework scritto in Python, capace di rendere uno script un web service disponibile su una porta. In particolare, lo script capace di effettuare le predizioni può essere chiamato tramite l'URL **localhost:5000/**.

L'API restituirà un JSON contenente la predizione tramite una response HTTP all'Adapter della piattaforma, il quale lo formatterà e permetterà all'apposito service di far visualizzare all'utente il risultato ottenuto.

In figura è mostrato lo schema rappresentante l'architettura ottenuta.





4.4 Risultati e conclusioni

La scelta del modello si è rivelata corretta dal punto di vista della sua semplicità e della corretta rappresentazione della realtà di riferimento. È stata però sottovalutata la gestione dei dati su cui effettuare il training. La maggior parte degli utenti ha risposto in maniera superficiale, e ci si è ritrovati con la scelta di alcune caratteristiche e di alcuni generi dominante rispetto ad altri. Ciò ha portato le risposte dell'agente a tendere verso i valori predominanti del dataset di training, ovvero ha causato overfitting.

Il risultato finale del modello è quindi una accuracy in media del 18%, calcolata grazie al 30% del dataset sul quale abbiamo effettuato delle predizioni, verificando successivamente la veridicità di queste ultime.

5 Sviluppi futuri

In questo paragrafo verranno discusse alcune idee per apportare, in futuro, miglioramenti all'agente intelligente attualmente proposto, con l'obiettivo di aumentare ulteriormente la qualità del servizio offerto al sistema BiblioNet.

1. migliorare la qualità di domande e risposte, in particolare per la fase di raccolta dei dati, in modo tale da avere un dataset più preciso e che non si specializzi su poche risposte;
2. aumentare il numero di campioni raccolti, puntando ad ottenere dati da diversi target, in modo che il dataset risulti più vicino possibile alla realtà di riferimento;
3. utilizzare tecniche migliorare e diversificare i dati di training, in quanto porterebbe a mitigare l'overfitting e ad aumentare conseguentemente l'accuracy;
4. l'idea sulla quale si fonda il modulo è quella di consigliare un genere letterario sulla base delle risposte ad un questionario, che nasce dall'ambiente BiblioNet ma può essere utilizzata anche a prescindere da esso. Una possibilità sarebbe, quindi, renderlo un servizio disponibile su un server, dato che è stato implementato come una API, in modo che possa essere sfruttato da tutti e reso disponibile online in qualsiasi forma.



6 Glossario

Termine	Significato
BiblioNet	Piattaforma web che permette un facile approccio al mando della lettura.
JSON	Tecnologia di formattazione dati.
Spring	Framework per piattaforme web.
flask	Micro-framework che rende uno script Python un web- service.
HTTP	Tecnologia di comunicazione tra applicazioni.
Adapter	Design pattern che permette ad oggetti con interfacce incompatibili di collaborare.
overfitting	Consiste nell'adattamento eccessivo di un modello statistico ai dati osservati.
dataset	Collezione di dati, ovvero i campioni raccolti.