

Documentazione progetto Ingegneria della conoscenza

Il progetto è stato realizzato da:

Salvatore Tucci (Matricola 723071)

Vincenzo Pasquadibisceglie (Matricola 723047)

Link github: https://github.com/VincenzoPasq/ICON_22.git

Il software nasce per aiutare le aziende informatiche nell'assunzione dei propri dipendenti fornendo loro un valore di probabilità di assunzione, in base alle conoscenze e all'esperienza del candidato.

La base di conoscenza è stata realizzata interamente dal gruppo, in quanto non è stato trovato un dataset che si avvicinasse molto all'idea di progetto scelta.

Requisiti Funzionali

Per eseguire correttamente il programma è necessario installare le seguenti librerie:

- **Pandas:** Per la manipolazione del Dataset
- **Sklearn:** Per eseguire gli algoritmi di Classificazioni
- **Statistics:** Per la misurazione dell'Accuratezza dei classificatori
- **Pytholog:** Per la realizzazione della Progettazione Logica
- **Owlready2:** Per l'implementazione dell'ontologia

MANUALE DELL'UTENTE

All'avvio del programma verrà visualizzato il seguente messaggio a linea di comando:

“Benvenuto nel sistema in grado di predire la probabilità di assunzione durante i colloqui con le Aziende. Di seguito verranno fatte domande inerenti alle abilità informatiche di cui si è in possesso.

Le risposte devono essere date nei possibili seguenti modi:

- Per il titolo di studio inserire:
 - > 0 (*NON DIPLOMATO*)
 - > 1 (*DIPLOMATO*)
 - > 2 (*LAUREATO*)
- Per i linguaggi informatici conosciuti utilizzare
 - > *sì, s*
 - > *no, n*
 - > *info* (per comprendere i requisiti del linguaggio richiesti)
- Per indicare l'esperienza lavorativa, indicare il numero di anni di lavoro nell'ambito informatico.”

In seguito, verranno mostrate le domande necessarie ad effettuare la predizione.

```
Di che titolo di studio sei in possesso?  
2  
Conosci il linguaggio Java?  
si  
Conosci il linguaggio MySql?  
si  
Conosci il linguaggio Python?  
si  
Conosci la lingua Inglese?  
no  
Quanti anni hai di esperienza lavorativa?  
0  
[{'Probabilità_Assunzione': '88%'}]  
  
Ottimo! Hai alte probabilità di essere richiamato per la candidatura all'interno di un'azienda!
```

Scelte progettuali

Nel nostro software si è scelto di utilizzare:

- Un *DATASET* interamente descritto dal gruppo per essere modellato in base alle esigenze di progetto.
- Un *CLASSIFICATORE* in grado di restituire la probabilità di essere assunto all'interno di un'azienda in base alle conoscenze che possiede l'utente interrogato. Inoltre, abbiamo utilizzato l'algoritmo *Leave-one-out cross validation* per calcolare l'accuratezza del classificatore.
- Il *PROLOG*, utilizzato per interrogare il KB verificando che rispetti determinate regole da cui verranno estratte le probabilità di essere assunti.
- Un *ONTOLOGIA*, realizzata per descrivere nel dettaglio il significato di conoscenza di un determinato linguaggio.

DATASET

Il programma è stato dotato di un *DATASET* per addestrare il classificatore. Le features contenute al suo interno si distinguono in:

- Titolo di Studio
- Conoscenze in Java
- Conoscenze in Python
- Conoscenze in MySql
- Conoscenze in inglese
- Esperienze Pregresse
- Possibilità di Assunzione

Ad ogni feature sono stati associati differenti domini di valori:

- Per il Titolo di studio, sono stati utilizzati valori interi che vanno da 0 a 2.
0 indica i non diplomati, 1 per i diplomati e 2 per i laureati;
- Per le conoscenze in Java, Python, MySql e Inglese, il valore booleano assegnato indica:
 - 0: non conoscenza del linguaggio
 - 1: conoscenza del linguaggio
- Per Esperienze Pregresse, un range di valori. I range possibili sono:
 - 0_2: da 0 a 2 anni di esperienza lavorativa
 - 3_6: da 3 a 6 anni di esperienza lavorativa
 - 7_ : 7 o più anni di esperienza lavorativa
- Per la possibilità di assunzione, i valori sono booleani.
 - 0: indica la possibilità di essere scartato
 - 1: indica la possibilità di essere assunto

CLASSIFICATORE

Abbiamo misurato l'accuratezza di vari classificatori optando per il Decision Tree Classifier, che ha ottenuto il livello di accuratezza più alto. La funzione accuracy ha permesso il calcolo di tale accuratezza, attraverso l'algoritmo 'LeaveOneOut CrossValidation'. La scelta di questo algoritmo ci è sembrata la più appropriata in quanto la dimensione del dataset è ridotta, essendo scritto a mano.

Modello	LeaveOneOut CV	Media
DecisionTree Classifier	0,909722	0,909722
DecisionTree Classifier	0,909722	
DecisionTree Classifier	0,909722	
DecisionTree Classifier	0,909722	
DecisionTree Classifier	0,909722	
RandomForest Classifier	0,861111	0,868056
RandomForest Classifier	0,854167	
RandomForest Classifier	0,861111	
RandomForest Classifier	0,888889	
RandomForest Classifier	0,875001	
GradientBoosting Classifier	0,916667	0,909722
GradientBoosting Classifier	0,909722	
GradientBoosting Classifier	0,902778	
GradientBoosting Classifier	0,916667	
GradientBoosting Classifier	0,902778	

Modello	LeaveOneOut CV	Media
GaussianNB	0,868056	0,868056
GaussianNB	0,868056	
GaussianNB	0,868056	
GaussianNB	0,868056	
GaussianNB	0,868056	
Kneighbors Classifier	0,895833	0,895833
Kneighbors Classifier	0,895833	
Kneighbors Classifier	0,895833	
Kneighbors Classifier	0,895833	
Kneighbors Classifier	0,895833	
ExtraTree Classifier	0,895833	0,895833
ExtraTree Classifier	0,895833	
ExtraTree Classifier	0,895833	
ExtraTree Classifier	0,895833	
ExtraTree Classifier	0,895833	

PROLOG

Abbiamo interrogato il KB composto da asserzioni che specificano le conoscenze dei vari linguaggi di un'utente, così da mostrare la percentuale di probabilità di essere assunto all'interno di un'azienda. In particolare, si riscontra una risposta negativa nel caso in cui le conoscenze attuali non siano sufficienti per essere assunto.

Per fare ciò abbiamo inserito in una KB, 2 proposizioni: lang e prob.

La prima permette di asserire la presenza delle conoscenze, a seconda delle risposte che l'utente fornisce; la seconda viene utilizzata in fase di interrogazione e restituirà la percentuale di assunzione.

```
kb([f"lang0(Java,{ris[0][0].lower()})",
    f"lang1(Python,{ris[0][1].lower()})",
    f"lang2(MySql,{ris[0][2].lower()})",
    f"lang3(Inglese,{ris[0][3].lower()})"],
```

```
"prob(X,Y,V) :- lang0(Java, X), lang1(Python, X), lang2(MySql, X), lang3(Inglese,X), V is '100%'",
"prob(X,Y,V) :- lang0(Java, Y), lang1(Python, Y), lang2(MySql, Y), lang3(Inglese,Y), V is '22%'",
"prob(X,Y,V) :- lang0(Java, X), lang1(Python, X), lang2(MySql, X), lang3(Inglese,Y), V is '88%'",
"prob(X,Y,V) :- lang0(Java, X), lang1(Python, X), lang2(MySql, Y), lang3(Inglese,X), V is '77%'",
```

ONTOLOGIA

L'introduzione dell'ontologia all'interno del progetto è stata utile per poter approfondire il significato di conoscenza di un determinato linguaggio. È stato utilizzato 'Protege', un software esterno, che ha permesso di rappresentare i vari linguaggi, fornendo loro una descrizione più specifica riguardo ciò che le aziende informatiche intendono, per conoscenza di un linguaggio. È possibile accedere alle descrizioni fornite dall'ontologia mediante il comando 'info', fornito all'utente nel momento in cui viene domandata la conoscenza di un linguaggio

```
Conosci il linguaggio Java?
info
Linguaggio: Java, descrizione: La conoscenza del linguaggio Java prevede:
- conoscere i principali pacchetti della libreria standard Java;
- apprendere ed utilizzare le collezioni Java (java.util);
- conoscere ed utilizzare il meccanismo di eccezioni Java;
- conoscere ed utilizzare il multi-threading in Java;
- apprendere la progettazione delle interfacce grafiche con componenti Swing (javax.swing);
- apprendere l'approccio Java per comunicazioni in rete e realizzare comunicazioni tra applicazioni in rete (java.net);
Conosci il linguaggio Java?
si
```

IMPLEMENTAZIONI FUTURE

In futuro, alcune feature che potrebbero essere implementate sono:

1. Inserimento GUI;
2. Fornire il nome dell'azienda in base alle conoscenze che l'utente possiede;

PROCESSO DI SVILUPPO E ORGANIZZAZIONE DEL LAVORO

Il progetto è stato realizzato completamente in remoto. La comunicazione tra i membri del team è avvenuta attraverso la piattaforma teams, che ha permesso la sincronizzazione e il supporto in tempo reale nella stesura del codice.

CONCLUSIONI

La realizzazione del progetto è stata molto impegnativa dal punto di vista della programmazione, in quanto non eravamo a piena conoscenza del linguaggio utilizzato e delle sue meccaniche di funzionamento. Questo primo approccio però, è stato fondamentale per migliorare e cimentarsi nell'applicazione dei concetti studiati durante il corso. Ci piacerebbe inoltre poter approfondire e migliorare il progetto, attraverso nuove implementazioni.

Grazie per l'attenzione.