

Sistema intelligente per la gestione del personale e creazione di terapie intensive

Documentazione Progetto

Gruppo di lavoro

- Andrea Mastrolonardo, 716698, a.mastrolonardo5@studenti.uniba.it
- Enrico Pappalardo, 727662, e.pappalardo1@studenti.uniba.it

<https://github.com/AMastro99/Progetto-Icon8-22-23>

AA 2022-2023

Introduzione

Il dominio di interesse da noi studiato e rappresentato in alcune delle sue parti è l'organizzazione di un ospedale di grandi dimensioni e la gestione del personale all'interno delle unità operative (U.O.) che lo costituiscono. A questo fine, ci siamo avvalsi della consulenza di un esperto del dominio, il quale ci ha fornito parametri e stime riguardanti le caratteristiche più rilevanti per l'obiettivo del progetto. I dati stipendiali sono stati invece estratti dai contratti collettivi nazionali di lavoro vigenti.

L'obiettivo del progetto è quello di studiare ed individuare le caratteristiche imprescindibili e preferenziali (avuto riguardo anche agli aspetti economico-finanziari), che il personale adibito ad una determinata unità operativa ospedaliera deve possedere per essere trasferito ad una terapia intensiva da creare all'interno della medesima unità. Più in particolare si è ipotizzata la possibilità che un'azienda ospedaliera abbia necessità di costituire al proprio interno e con le risorse umane che già dispone una struttura di cure intensive o semi-intensive; ad esempio, di istituire la Rianimazione all'interno della U.O. di Anestesia o un ambulatorio semi-intensivo di Dialisi nell'ambito della U.O. di Nefrologia. Riteniamo di dover precisare subito che la differenza fra una terapia intensiva ed un'altra semi-intensiva consiste fondamentalmente nell'assistenza che nel primo caso deve essere garantita 24 ore su 24.

Per questo motivo abbiamo preso in considerazione non tutto il personale di un ospedale di grandi dimensioni, ma soltanto quello in dotazione a sei unità operative complesse (unità di rilevante importanza strategica per l'azienda), all'interno delle quali poter istituire delle unità di terapia intensiva/semi-intensiva.

Creazione Dataset e Scelte di progetto

Per meglio effettuare uno studio sul dominio di interesse da noi scelto, abbiamo creato, seguendo i consigli e indicazioni fornite da un esperto del dominio da noi consultato, un dataset contenente 500 "dipendenti" fittizi di un ospedale ipotetico in sede a Bari composto da 6 unità operative complesse (reparti) elencate di seguito: Cardiologia, Anestesia e Rianimazione, Neurologia, Nefrologia e Dialisi, Malattie dell'apparato respiratorio e Neonatologia.

Ciascun elemento appartenente al dataset è suddiviso in più campi:

- Matricola (Codice identificativo per individuare univocamente ciascun dipendente);
- Nome (Nome del dipendente)
- Cognome (Cognome del dipendente)
- Data di nascita (Data di nascita del dipendente)
- Sesso (Sesso del dipendente)
- Numero di Cellulare (Numero di cellulare del dipendente, utilizzato per contattarlo casi di emergenza o urgenza)
- Residenza (Indirizzo di residenza del dipendente, contiene la via, il numero civico, il tipo di abitazione, il codice di avviamento postale CAP, la città e la provincia)
- Qualifica (Nome della figura professionale a cui appartiene il dipendente)
- Titolo (Titolo abilitante conseguito dal dipendente)
- Limitazioni (Booleano che indica se il dipendente usufruisce dei benefici forniti dalla legge 104 del 1992)
- Reparto (Nome dell'unità operativa complessa a cui è assegnato il dipendente)
- Specializzazione (Nome della specializzazione conseguita dal dipendente infermieristico)

- Anzianità (Numero di anni di servizio del dipendente)
- Classificazione (Nome della classificazione associata al personale medico)
- Stipendio base (Cifra dello stipendio del dipendente non considerando gli incrementi relativi all'anzianità e alla classificazione del personale medico)
- Incremento Anzianità (Cifra dell'incremento legato all'anzianità lavorativa del dipendente)
- Bonus Incarico (Cifra dell'incremento legato alla classificazione del personale medico)
- Stipendio Totale (Cifra ottenuta dalla somma dello stipendio base più i vari incrementi relativi all'anzianità lavorativa e alla classificazione del personale medico)

Per la creazione del suddetto dataset abbiamo scritto uno script in Python con l'obiettivo di popolare un file **dataset.csv** con 500 tuple (+1 per i nominativi dei campi), ciascuna rappresentante un dipendente dell'ospedale da noi ipotizzato. Il numero dei dipendenti è stato scelto affinché ci sia abbastanza personale per creare una terapia intensiva per ciascuna unità operativa presente nell'ospedale (6 nel nostro caso).

Per la matricola associata al dipendente abbiamo deciso di utilizzare la seguente struttura: A00001, ovvero la lettera A seguita da 5 cifre che rappresentano il numero associato al dipendente.

Per l'anagrafica di ciascun dipendente abbiamo utilizzato la libreria Faker [1] di Python, la quale genera tutti i dati relativi, come nome, cognome, data di nascita, numero di cellulare e residenza.

Per alcuni di questi campi, come la data di nascita, il numero di cellulare e residenza sono stati imposti dei vincoli durante la creazione o durante la selezione.

A parte questi campi, abbiamo deciso di includere anche il sesso del dipendente; per questo campo abbiamo deciso di utilizzare la funzione `gender_guesser()` dalla libreria Detector, la quale ha il compito di assegnare il sesso di un dipendente analizzando il nome già creato dalla libreria Faker.

Per la data di nascita abbiamo scelto di includere come età del personale minima il valore di 24 e come età del personale massima il valore di 67; questa scelta è stata presa dopo aver consultato il nostro esperto del dominio, che ci ha indicato i valori estremi del personale che di solito si assume nel contesto da noi ipotizzato.

Per il numero di cellulare abbiamo scelto di creare solamente numeri con il prefisso italiano +39.

Per il campo residenza abbiamo scelto di inserire nel dataset solamente le residenze create incluse nella provincia di Bari; questa scelta è stata presa ipotizzando che, se un dipendente venisse assunto in una qualsiasi azienda in Bari, allora è sensato presumere che egli abbia residenza nelle vicinanze di Bari.

Una volta creata l'anagrafica del dipendente, lo script inizia a popolare i campi relativi alle sue competenze e al ruolo che ricopre all'interno dell'organizzazione dell'ospedale.

I valori possibili all'interno del campo qualifica e le rispettive percentuali di scelta sono le seguenti: caposala (uno per ciascuna unità operativa presente nell'ospedale, quindi, in totale 6), ausiliario (9.3%), OSS (9.3%), infermiere (57.1%) e medico (24.3%). Per ciascuno di questi valori è stata indicata una percentuale di assegnazione in modo tale che lo script scelga a random, seguendo le probabilità, uno dei possibili valori.

Per il campo titolo, abbiamo deciso di far in modo che lo script assegni i valori possibili del campo a seconda del valore presente nel campo qualifica. I valori possibili per il campo titolo sono i seguenti: no, diploma operatore sociosanitario, laurea scienze infermieristiche, laurea magistrale infermieristica e laurea medicina. Lo script sceglierà dunque il valore nel campo titolo che si suppone il dipendente debba conseguire per ottenere il valore presente nel campo qualifica.

Successivamente il programma assegna il valore del campo limitazioni lavorative; essendo un valore booleano, abbiamo scelto un 20% di assegnare il valore 1 e 80% di assegnare il valore 0. In questo caso la scelta non si basa su nessun valore in quanto si tratta di una caratteristica personale indipendente dalla professione di ciascun dipendente.

Il prossimo campo che lo script popola è il campo reparto. Per questo campo il programma osserva il valore presente all'interno di qualifica e, seguendo anche le probabilità da noi assegnate, prosegue con la scelta del valore relativo al campo reparto.

I valori possibili all'interno del campo reparto e le relative probabilità sono le seguenti: cardiologia (23%), anestesia e rianimazione (25%), nefrologia e dialisi (10%), neurologia (17%), malattie dell'apparato respiratorio (13%), neonatologia (12%). Come già indicato in precedenza, i valori riguardanti le percentuali sono stati scelti consultando il nostro esperto del dominio in modo che rappresentino una distribuzione verosimile del personale in rapporto alle unità operative presenti. Per la scelta del valore all'interno del campo specializzazione, lo script controlla che il valore all'interno del campo qualifica sia uguale a medico e basa la scelta secondo il reparto assegnato al dipendente e una distribuzione di probabilità da noi fornita.

I valori possibili all'interno del campo specializzazione e le rispettive probabilità di scelta sono: se il reparto è "cardiologia", allora al 50% viene scelto un valore appartenente alla lista di specializzazioni in parentesi ("Cardioangiopatie", "Cardiologia e reumatologia", "Cardiologia pediatrica", "Fisiopatologia cardiocircolatoria/ cardiovascolare", "Malattie cardiovascolari", "Malattie cardiovascolari e reumatiche", "Semeiotica cardiovascolare", "Allergologia e immunologia clinica", "Angiologia", "Geriatria", "Malattie metaboliche e Diabetologia", "Malattie dell'apparato respiratorio", "Medicina e Chirurgia d'accettazione e d'urgenza", "Medicina Interna", "Medicina dello Sport", "Reumatologia") e al restante 50% viene scelta la specializzazione "Cardiologia e malattie dei vasi";

se il reparto è "anestesia e rianimazione", allora al 50% viene scelto un valore appartenente alla lista di specializzazioni in parentesi ("Anestesia", "Anestesiologia e rianimazione", "Anestesiologia generale e speciale odontostomatologica", "Anestesiologia rianimazione e terapia intensiva", "Medicina subacquea ed iperbarica", "Servizio sanitario d'urgenza ed emergenza", "Terapia del dolore", "Terapia intensiva"), oppure al 50% viene scelta la specializzazione "Anestesiologia";

se il reparto è "nefrologia e dialisi", allora al 50% viene scelto uno tra i valori appartenenti alla lista ("Emodialisi", "Malattie del rene del sangue e del ricambio", "Nefrologia e dialisi", "Nefrologia pediatrica", "Allergologia ed Immunologia Clinica Geriatria", "Medicina Interna", "Urologia") e al restante 50% viene scelta la specializzazione "Nefrologia medica";

se il reparto è "neurologia", allora al 50% viene scelta una specializzazione appartenente alla lista (Clinica neurologica, "Malattie nervose", "Malattie nervose e mentali", "Neurologia d'urgenza", "Neurologia e psichiatria", "Neuropsichiatria", "Neuroriabilitazione", "Semeiotica neurologica", "Terapia neurologica", "Medicina fisica e della riabilitazione", "Neurofisiologia clinica", "Neurofisiopatologia", "Neuropatologia", "Neuropsichiatria infantile", "Neuroradiologia", "Psichiatria") oppure al 50% viene scelta la specializzazione "Neurofisiologia clinica";

se il reparto è "malattie dell'apparato respiratorio", allora al 50% viene scelto uno tra i valori appartenenti alla lista ("Fisiopatologia e fisiochinesiterapia respiratoria", "Fisiopatologia respiratoria", "Malattie dell'apparato respiratorio", "Tisiologia", "Allergologia ed Immunologia Clinica", "Cardiologia", "Geriatria", "Malattie Infettive", "Medicina e Chirurgia di Accettazione ed Urgenza", "Medicina Interna", "Medicina dello Sport", "Oncologia") e al restante 50% viene scelta la specializzazione "Pneumologia";

se il reparto è "neonatologia", allora al 50% viene scelta una specializzazione appartenente alla lista ("Clinica pediatrica", "Pediatria", "Pediatria e puericultura", "Pediatria preventiva e puericultura", "Pediatria preventiva e sociale", "Pediatria sociale e puericultura", "Puericultura", "Puericultura e dietetica infantile", "Puericultura ed igiene infantile", "Puericultura dietetica infantile ed assistenza sociale dell'infanzia", "Terapia intensiva neonatale", "Pediatria") oppure al 50% viene scelta la specializzazione "Patologia neonatale".

Per ogni dipendente che non ha valore di medico all'interno del campo qualifica, il valore di

specializzazione che gli verrà assegnato sarà no.

Assegnata la specializzazione, lo script passa al calcolo dell'anzianità di servizio calcolando gli anni del dipendente e sottraendo 23. Nel progetto abbiamo deciso di supporre 23 come l'età che ogni dipendente aveva quando sono stati assunti dall'ospedale.

Per la scelta del valore nel campo classificazione il programma controlla che il dipendente abbia come qualifica il valore di medico; se l'esito del controllo è positivo allora si passa alla scelta del valore seguendo una distribuzione di probabilità da noi indicata.

Prima lo script assegna 6 volte il valore di "incarico di responsabile di struttura complessa", una volta per un medico in ciascuna unità operativa presenta nell'ospedale (nel nostro caso 6 reparti).

Una volta eseguita questa assegnazione, per ogni medico successivo si seguirà il seguente criterio: all'85% assegnare il valore di "incarico professionale" e al 15% assegnare il valore di "incarico di responsabile di struttura semplice".

Per ogni dipendente che ha esito negativo nel controllo, e quindi non ha medico come valore di qualifica, lo script assegnerà il valore di "no".

Successivamente lo script passa alla creazione dei valori da inserire nei campi relativi allo stipendio del dipendente.

Per il campo stipendio base, lo script esegue l'assegnazione dei valori una volta letto il contenuto del campo qualifica.

Se il dipendente ha come valore di qualifica ausiliario, allora avrà come stipendio base (annuale) il valore di 20907€;

se il valore di qualifica è OSS, allora lo stipendio base (annuale) sarà di 22180€;

se il valore di qualifica è infermiere, allora lo stipendio base (annuale) sarà di 23653€;

se il valore di qualifica è caposala, allora lo stipendio base (annuale) sarà di 25675€;

se il valore di qualifica è medico, allora lo stipendio base (annuale) sarà di 45260€.

Per quanto riguarda il campo incremento anzianità esso viene assegnato a tutti i dipendenti con qualifiche diverse da medico e sarà espresso in euro in più all'anno rispetto allo stipendio base.

Se la qualifica del dipendente è ausiliario, allora gli scatti relativi all'anzianità lavorativa sono i seguenti: 0€ se anzianità < 5, 700€ se 5 < anzianità < 10, 1400€ se 10 < anzianità < 15, 2100€ se 15 < anzianità < 20, 2800€ se 20 < anzianità < 25, 3500€ se 25 < anzianità < 30, 4200€ se anzianità >= 30.

Se la qualifica del dipendente è OSS, allora gli scatti relativi all'anzianità lavorativa sono i seguenti: 0€ se anzianità < 5, 800€ se 5 < anzianità < 10, 1600€ se 10 < anzianità < 15, 2400€ se 15 < anzianità < 20, 3200€ se 20 < anzianità < 25, 4000€ se 25 < anzianità < 30, 4800€ se anzianità >= 30.

Se la qualifica del dipendente è infermiere, allora gli scatti relativi all'anzianità lavorativa sono i seguenti: 0€ se anzianità < 5, 1000€ se 5 < anzianità < 10, 2000€ se 10 < anzianità < 15, 3000€ se 15 < anzianità < 20, 4000€ se 20 < anzianità < 25, 5000€ se 25 < anzianità < 30, 6000€ se anzianità >= 30.

Se la qualifica del dipendente è caposala, allora gli scatti relativi all'anzianità lavorativa sono i seguenti: 0€ se anzianità < 5, 1200€ se 5 < anzianità < 10, 2400€ se 10 < anzianità < 15, 3600€ se 15 < anzianità < 20, 4800€ se 20 < anzianità < 25, 6000€ se 25 < anzianità < 30, 7200€ se 30 < anzianità < 35, 8400€ se anzianità >= 35.

Per i dipendenti con qualifica di medico il programma assegnerà 0 come valore del campo incremento anzianità.

Per il campo bonus incarico, lo script popolerà solamente le tuple appartenenti a dipendenti con qualifica di medico e sceglierà il valore da inserire a seconda del valore all'interno del campo classificazione.

Se il dipendente ha come classificazione il valore incarico professionale, allora il suo bonus incarico sarà di 24143€ (annuali).

Se il dipendente ha all'interno di classificazione il valore incarico di responsabile di struttura semplice, allora il suo bonus incarico sarà di 33333€ (annuali).

Se il dipendente ha come valore di classificazione incarico di responsabile di struttura complessa, allora il suo bonus incarico sarà di 45702€ (annuali).

Per tutti i dipendenti che non hanno medico come valore del campo qualifica, lo script inserirà il valore 0 all'interno del campo bonus incarico.

Infine, il campo stipendio totale sarà popolato eseguendo una somma dei valori all'interno dei campi stipendio base, incremento anzianità e bonus incarico (stipendio totale = stipendio base + incremento anzianità + bonus incarico).

Una volta eseguita la creazione e popolazione di 500 tuple, il dataset verrà salvato all'interno di un file denominato "**dataset.csv**". Durante l'esecuzione delle varie funzionalità del sistema da noi creato verranno create copie del dataset con campi rimossi o aggiunti a seconda delle necessità del sistema.

È necessario annotare che il dataset è stato modificato durante la nostra fase di verifica delle funzionalità da noi implementate affinché sia presente un numero di personale congruo alle aspettative massime del sistema, soprattutto durante la ricerca e assegnazione del personale alle terapie intensive.

Elenco argomenti di interesse

- Knowledge Base (KB)
- Creazione e gestione terapie intensive (CSP)
- Apprendimento non supervisionato (K-Means)

Knowledge Base (KB)

Il nostro sistema intelligente è diviso in diversi moduli, ognuno dei quali svolge determinati compiti e funzioni per permettere uno studio efficace del personale e le caratteristiche ad esso associate anche all'interno delle diverse terapie intensive.

Il Knowledge Base System è stato implementato tramite l'utilizzo del linguaggio Prolog e ha l'obiettivo di rappresentare la conoscenza necessaria alla rappresentazione del personale di un ospedale e, in particolare per il nostro progetto, la gestione e assegnazione di questo personale all'interno di diverse terapie intensive, ciascuna legata ad un determinato reparto già presente all'interno del mondo ideale che vogliamo descrivere.

Per la realizzazione e funzionamento della Knowledge Base è stato necessario installare la versione 8.4.2 di SWI-Prolog, importare la libreria Pyswip [2] e scaricare, attraverso il terminale dell'IDE, i pacchetti necessari al funzionamento della suddetta libreria tramite il comando "pip install git+https://github.com/ML-KULEuven/pyswip".

Successivamente abbiamo definito e creato gli assiomi, regole e query con la quale l'utente potrà interrogare la Knowledge Base ed interfacciarsi con essa.

Al primo avvio del programma verrà generato il file "KB.pl" che conterrà tutte le informazioni citate precedentemente, ottenute tramite la conoscenza di fondo contenuta all'interno del dataset da noi costruito.

Interfacciarsi e interrogare la KB

Avviato il programma, l'utente potrà, una volta eseguito l'accesso all'apposita area, interrogare la KB tramite l'utilizzo di query da noi implementate e riportate di seguito:

1. Massimo e minimo dello stipendio in base alla qualifica: in input la query richiederà una delle qualifiche presenti all'interno della Knowledge Base (ausiliario, OSS, infermiere, e medico). Una volta ottenuta la qualifica, la query interrogherà la KB, restituendo come input il salario massimo e minimo relativo alla qualifica indicata;

es.

```
===== SALARIO MINIMO E MASSIMO PER QUALIFICA =====
1. ausiliare
2. OSS
3. infermiere
4. caposala
5. medico
0. Annulla operazione
Scegli la qualifica per la quale si vuole stampare il salario minimo e massimo:
3

Il salario minimo per la qualifica di "infermiere" è 23653.00 euro.
Il salario massimo per la qualifica di "infermiere" è 29653.00 euro.
```

2. Visualizza la regola con la quale viene calcolato lo stipendio: in input la query richiederà una delle qualifiche presenti all'interno della KB e restituirà come output la regola utilizzata per il calcolo dello stipendio, eseguito, a seconda della qualifica indicata, in base agli anni di anzianità o, nel caso dei medici, in base alla classificazione;

es.

```
===== REGOLA PER IL CALCOLO DEL SALARIO PER QUALIFICA =====
1. ausiliare
2. OSS
3. infermiere
4. caposala
5. medico
0. Annulla operazione
Scegli la qualifica per la quale si vuole mostrare la regola per il calcolo del salario:
5

Stipendio totale = stipendio base + bonus incarico

Lo stipendio base per la qualifica di medico è: 45260 euro.
Bonus incarico:
Un medico con un 'incarico professionale' riceverà un bonus incarico di 24143 euro, per un totale di 69403 euro.
Un medico con un 'incarico di responsabile di struttura semplice' riceverà un bonus incarico di 33333 euro, per un totale di 78593 euro.
Un medico con un 'incarico di responsabile di struttura complessa' riceverà un bonus incarico di 45702 euro, per un totale di 90962 euro.
```

3. Visualizza in ordine decrescente le qualifiche in base allo stipendio: questa query interrogherà la KB analizzando tutti gli stipendi base relativi alle qualifiche all'interno della conoscenza. Terminata l'interrogazione, la query restituirà come output una classifica delle qualifiche più remunerative, ordinate in maniera decrescente, sempre in base allo stipendio base (in quanto lo stipendio totale prende in considerazione caratteristiche individuali del personale, come l'anzianità);

es.

```
===== QUALIFICHE PIU' REMUNERATIVE =====
Le qualifiche più remunerative in ordine decrescente sono:
1. medico con incarico di responsabile di struttura complessa
2. medico con incarico di responsabile di struttura semplice
3. medico con incarico professionale
4. caposala
5. infermiere
6. OSS
7. ausiliare
```

4. Media degli stipendi in base alla qualifica: tramite questa query, l'utente potrà chiedere alla KB, inserendo in input una delle qualifiche presenti all'interno della KB, di restituire lo stipendio medio relativo alla qualifica inserita. Lo stipendio medio, in questo caso, verrà calcolato considerando lo stipendio totale relativo al singolo dipendente invece dello stipendio base in quanto le variazioni sullo stipendio base sono di natura individuale;

es.

```
===== SALARIO MEDIO PER QUALIFICA =====
1. ausiliare
2. OSS
3. infermiere
4. caposala
5. medico
0. Annulla operazione
Scegli la qualifica per la quale si vuole calcolare lo stipendio medio:
4

Il salario medio per la qualifica di "caposala" è 30075.00 euro
```

5. Confronto di due qualifiche in base agli stipendi: questa query richiederà in input due qualifiche presenti all'interno della KB. La query calcolerà quindi gli stipendi medi relativi alle due qualifiche indicate e restituirà come output qual è la relazione tra gli stipendi medi calcolati, esplicitando se la prima qualifica inserita è più o meno remunerativa della seconda qualifica inserita. Prima del loro

inserimento, all'utente sarà chiesto se il confronto deve tener conto dell'anzianità e della classificazione o meno (questa distinzione viene eseguita per individuare se è necessario utilizzare lo stipendio base o lo stipendio totale durante il calcolo dello stipendio medio). Qualora l'utente non volesse considerare i campi anzianità e classificazione, allora la query restituirà come output semplicemente la relazione tra gli stipendi base delle due qualifiche indicate;

es.

```
===== CONFRONTO FRA QUALIFICHE =====
Desideri comparare gli stipendi tenendo conto dell'anzianità e della classificazione dei medici?
1. Si
2. No
1

Scegli la prima qualifica da confrontare tra le seguenti:
1. ausiliare
2. OSS
3. infermiere
4. caposala
5. medico con incarico professionale
6. medico con incarico di responsabile di struttura semplice
7. medico con incarico di responsabile di struttura complessa
7

Scegli la seconda qualifica da confrontare tra le seguenti:
1. ausiliare
2. OSS
3. infermiere
4. caposala
5. medico con incarico professionale
6. medico con incarico di responsabile di struttura semplice
2

Il salario della qualifica di medico con incarico di responsabile di struttura complessa è maggiore del salario medio della qualifica di "OSS" di 66032.00 euro.
Il salario per la qualifica di medico con incarico di responsabile di struttura complessa è 90962.00 euro.
Il salario medio per la qualifica di OSS è 24930.00 euro, con un'anzianità media di 20 anni.
```

6. Visualizza la graduatoria del personale con lo stipendio più alto rispetto alla qualifica: questa query richiederà come input la qualifica sulla quale la KB dovrà essere interrogata. La query restituirà come output i primi dieci dipendenti con la stessa qualifica indicata in input in graduatoria, ordinati in base al loro stipendio totale. La graduatoria ottenuta come output sarà in ordine decrescente;

es.

```
===== CLASSIFICA DEL PERSONALE PER STIPENDIO =====
1. ausiliare
2. OSS
3. infermiere
4. caposala
5. medico
0. Annulla operazione
Scegli la qualifica per la quale si vogliono determinare la graduatoria in base allo stipendio:
4

Matricola   Nome      Cognome    Anzianità   Salario totale
-----
'A00131'    'Eva'     'Pascarella'  37         34075
'A00497'    'Enrico'  'Vianello'   42         34075
'A00181'    'Maurilio' 'Boccherini' 29         31675
'A00008'    'Lucia'   'Abbagnale'  12         28075
'A00396'    'Elena'   'Rossini'    6          26875
'A00320'    'Valeria' 'Bersani'     2          25675
```

7. Visualizza la graduatoria del personale con lo stipendio più basso rispetto alla qualifica: questa query richiederà come input la qualifica sulla quale la KB dovrà essere interrogata. La query restituirà come output i primi dieci dipendenti con la stessa qualifica indicata in input in graduatoria, ordinati in base al loro stipendio totale l'ordine della graduatoria ottenuta come output sarà crescente;

es.

```

===== CLASSIFICA DEL PERSONALE PER STIPENDIO =====
1. ausiliare
2. OSS
3. infermiere
4. caposala
5. medico
0. Annulla operazione
Scegli la qualifica per la quale si vogliono determinare la graduatoria in base allo stipendio:
1

Matricola   Nome      Cognome    Anzianità   Salario totale
-----
'A00030'    'Fabio'    'Molcini'    1           20907
'A00240'    'Etta'     'Salgari'    1           20907
'A00378'    'Ottavio'  'BOLDU'      4           20907
'A00100'    'Annalisa' 'Palazzo'    6           21607
'A00276'    'Massimo'  'Ciampi'     7           21607
'A00108'    'Lara'     'Trillini'   12          22307
'A00127'    'Cassandra' 'Chiappetta' 13          22307
'A00355'    'Carla'    'Casadei'    11          22307
'A00399'    'Bernardo' 'Mondaini'   11          22307
'A00417'    'Fiorenzo' 'Curci'     10          22307

```

8. Visualizza caratteristiche del dipendente tramite la sua matricola: questa query richiederà come input la matricola di un dipendente presente nel dataset. Una volta interrogata la KB, la query restituirà come output la scheda anagrafica, la quale includerà i dati anagrafici e le sue caratteristiche più importanti relative al ruolo che assume all'interno dell'ospedale;

es.

```

===== TROVA PERSONALE DALLA MATRICOLA =====
Inserisci la matricola per cercare le informazioni del personale:
A00274

matricola   nome      cognome    qualifica    anzianità    classificazione    stipendio base    incremento anzianità    bonus incarico    stipendio totale
-----
A00274      Pietro    Tamborini  OSS          19          no                22180            2400                  0                24580

```

9. Visualizza la suddivisione del personale nei cluster: questa query non necessita di alcun valore in input e restituisce come output la visualizzazione dei raggruppamenti del personale durante l'esecuzione del modulo di clustering. Per ogni cluster (nel nostro caso 2) sarà indicato il numero dei dipendenti ad esso assegnati.

es.

```

===== VISUALIZZA NUMERO DI CLUSTER =====
Personale appartenente al cluster 0: 369
Personale appartenente al cluster 1: 131

```

10. Visualizza il personale assegnato ad una terapia intensiva: questa query richiede come input il nome associato alla terapia intensiva della quale si vuole conoscere l'organizzazione. Una volta interrogata la KB, la query restituirà come output le qualifiche presenti all'interno della terapia intensiva indicata e, per ciascuna qualifica, le matricole dei dipendenti assegnati ad essa. Qualora venga indicata in input una terapia intensiva non ancora creata, la query restituirà come output un messaggio che indicherà di non aver trovato alcun risultato.

```

===== VISUALIZZA ORGANIZZAZIONE DI UNA TERAPIA INTENSIVA =====
1. utic
2. rianimazione
3. stroke_unit
4. dialisi
5. respiratoria
6. utin
0. Annulla operazione
Seleziona la terapia intensiva:
4

Organizzazione del personale:
infermiere: 10 -> "A00005", "A00070", "A00144", "A00211", "A00244", "A00307", "A00407", "A00441", "A00495", "A00496";
caposala: 1 -> "A00131";
medico: 4 -> "A00029", "A00060", "A00275", "A00498";

```

CSP – Constraint Satisfaction Problems

Sommario

L'obiettivo di questo modulo della Knowledge Base è permettere l'istituzione di terapie intensive da parte dell'utente e la selezione del personale da assegnare ad esse a seconda dei loro profili professionali, della loro funzionalità e anche del loro costo economico, il tutto parametrato alle dimensioni della terapia intensiva da istituire. Per questo motivo si è partiti solo dal personale assegnato alle Unità Operative Complesse di seguito elencate, ipotizzando la creazione di strutture intensive o semi-intensive a fianco delle stesse precisate:

Cardiologia – Unità Terapia Intensiva Coronarica (UTIC)

Anestesia e Rianimazione – Rianimazione e Rianimazione Post-Operatoria

Neurologia – Stroke Unit

Nefrologia e Dialisi – Dialisi

Pneumologia – Terapia Intensiva Respiratoria

Neonatologia – Unità Terapia Intensiva Neonatale (UTIN)

Appare ovvio che l'ipotesi di mondo da noi considerata è condizionata da numerose variabili, prime fra le quali le autorizzazioni regionali, i finanziamenti pubblici e il numero dei posti letto da attivare. Per conseguenza, qualora il numero dei posti letto da istituire risultasse alto (nella nostra ipotesi fino ad un massimo di 10), la terapia intensiva dovrebbe essere dotata di un caposala e magari anche di un direttore medico di struttura complessa, figure professionali che dovrebbero risultare superflue in ipotesi di terapia intensiva con limitato numero di posti letto.

Per la creazione di ciascuna terapia intensiva, all'utente verrà chiesto di eseguire alcune scelte al fine di definire vincoli di ricerca a partire da esse. Per prima cosa verrà chiesto di inserire il numero di posti letto disponibili e in seguito verrà richiesto di inserire il budget, il quale servirà da vincolo economico-finanziario durante la scelta del personale; se il valore del budget è minore del costo minimo da noi calcolato per ciascuna terapia intensiva, allora il programma fornirà un messaggio di errore, chiedendo di inserire un valore congruo.

Anche se non rientra negli scopi di questo progetto, riteniamo di precisare che, quando parliamo di budget intendiamo la spesa del personale della singola terapia intensiva da istituire; questa non costituisce una spesa aggiuntiva, supplementare in quanto nelle premesse abbiamo già precisato che l'apertura della terapia intensiva avverrà utilizzando del personale già assegnato all'unità operativa complessa "madre", cioè la principale. Pertanto, le risorse dei fondi regionali o pubblici necessari e decisivi per la quantificazione dei posti letto saranno utilizzate per finanziare strumentazione ed apparecchiature che, in alcuni casi (ad esempio il rene artificiale per la Dialisi) possono essere anche estremamente costose. In ogni caso abbiamo voluto tener conto del budget per monitorare la spesa della terapia intensiva quale dato comunque imprescindibile.

```
===== Scelta reparti di terapia intensiva =====
1. UNITA' TERAPIA INTENSIVA CORONARICA (UTIC)
2. RIANIMAZIONE E RIANIMAZIONE POST-OPERATORIA
3. STROKE-UNIT
4. DIALISI
5. TERAPIA INTENSIVA RESPIRATORIA
6. UNITA' TERAPIA INTENSIVA NEONATALE (UTIN)
0. Torna indietro.
Inserire il numero corrispondente al reparto che si desidera creare:
6
Inizio processo di apertura del reparto di terapia intensiva UTIN
Inserire il numero di posti letto disponibili all'interno della terapia intensiva:
9
Inserire il budget disponibile per la creazione della terapia intensiva (esclusi oneri):
900000
```

Le successive scelte riguardano i criteri di preferenza legati alle caratteristiche del personale, in particolare la preferenza all'inserimento di infermieri giovani o in base all'esperienza maturata in servizio e la preferenza all'inserimento di medici con incarico di struttura semplice o struttura complessa, scelta a seconda della valenza strategica che l'utente intende assegnare alla struttura di terapia intensiva o semi-intensiva che sta istituendo.

```
Scegliere se si preferisce inserire infermieri giovani o in base all'esperienza maturata in servizio.

===== Preferenza anzianita' di servizio degli infermieri =====
1. Infermieri sotto i 20 anni di anzianita'
2. Infermieri uguali o superiori a 20 anni di anzianita'
0. Torna indietro
Inserire il numero corrispondente alla preferenza da scegliere:
1
Selezionato infermieri sotto i 20 anni di anzianita'

Scegliere se si vuole inserire un medico responsabile di struttura semplice o di struttura complessa a seconda della valenza strategica che si intende assegnare alla struttura.

===== Preferenza classificazione medico responsabile =====
1. Medico con classificazione di responsabile di struttura semplice
2. Medico con classificazione di responsabile di struttura complessa
0. Torna indietro
Inserire il numero corrispondente alla preferenza da scegliere:
2
Selezionato medico con classificazione di responsabile di struttura complessa
```

Ottenuti questi dati, il sistema fornisce una dotazione complessiva del personale medico e paramedico, basata sul numero di posti letto inseriti precedentemente dall'utente.

Una volta completata questa fase, il sistema inizia a creare il problema CSP relativo alla terapia intensiva scelta. Questo include la dichiarazione delle variabili, i loro relativi domini, e i vincoli che verranno posti su di essi. Per la maggior parte delle terapie intensive, le variabili considerate sono le seguenti, con relativi domini all'interno delle parentesi: Qualifica (di cui ci interessano solo i valori OSS, infermiere, caposala e medico), Classificazione (con i valori di incarico professionale, incarico di responsabile di struttura semplice e incarico di responsabile di struttura complessa), Specializzazione (i cui valori cambiano a seconda dell'Unità Operativa nel rispetto di quanto stabilito dai D.P.R. 483 e 484 del 1997 e suoi successivi aggiornamenti), Anzianità (dal valore minimo di 0 ad un valore massimo di 43), Limitazioni lavorative (un booleano che indica se il dipendente usufruisce o meno dei benefici della legge 104/92), Sesso (un booleano che indica se il sesso alla nascita del dipendente è donna o uomo), Titolo (i cui valori possono essere: diploma socio-sanitario, laurea scienze infermieristiche, laurea magistrale infermieristica e laurea medicina) e residenza (il cui valore intero include la strada, il civico, il palazzo, il piano, il CAP, la città e la provincia ma, per semplificare l'analisi, con lo stesso grado di accuratezza, abbiamo preso in considerazione solo il CAP).

La dichiarazione dei vincoli dipende sia dalla scelta della terapia intensiva da voler istituire sia dal ruolo di cui si sta cercando il personale.

Alcuni dei vincoli particolari imposti a seconda dei ruoli da ricercare sono i seguenti:

- Per gli OSS: valore del campo "limitazioni lavorative" uguale a zero;
- Per gli Infermieri: valore del campo "residenza" appartenente ad una città limitrofa a Bari;
- Per i Caposala: valore del campo "titolo" uguale a "laurea magistrale infermieristica";
- Per i Medici: valore del campo "specializzazione" appartenente ad una lista di specializzazioni definita da noi in quanto inerenti a ciascuna terapia intensiva;
- Per i Responsabili: oltre al vincolo posto ai medici, il valore del campo "anzianità" deve essere superiore a cinque.

Ogni problema CSP verrà risolto con una serie di ricerche, ognuna delle quali troverà il numero del personale per ciascun ruolo all'interno della terapia intensiva selezionata. Effettuata la ricerca per un determinato ruolo, il CSP verrà dichiarato nuovamente in modo da inserire i vincoli riguardanti il ruolo da ricercare successivamente (motivazione di questa procedura in seguito).

Alla fine di ogni ricerca, il personale selezionato verrà segnalato come assegnato e le matricole relative a ciascun dipendente verranno inserite in una lista, la quale verrà salvata in un file specifico (indicato come “***personale_nometerapiaintensiva***”) e rappresenterà il personale assegnato alla terapia intensiva per la quale si è effettuato il processo di ricerca.

Per visualizzare la lista del personale relativa ad una terapia intensiva già istituita, l’utente potrà utilizzare il comando apposito selezionandolo tra le opzioni all’interno del menu principale del modulo dedicato alle terapie intensive, questo comando effettuerà, selezionata una terapia intensiva precedentemente creata con successo, una ricerca all’interno del dataset per ciascuna matricola presente nella lista del personale relativo alla terapia intensiva indicata; terminata la ricerca, il comando restituirà come output una lista, ordinata secondo il valore della matricola, composta da ciascun dipendente appartenente alla terapia intensiva creata e tutte le informazioni disponibili per ciascuno di loro.

Matricola	Nome	Cognome	Data di nascita	Sesso	Numero di cellulare	Residenza	Qualifica	Titolo
A00003	Romana	Iacovelli	1986-05-01	0	+39 3482859351	Viale Micheletti, 79 70129, Ceglie Del Campo (BA)	Infermiere	laurea scienze infermieristiche
A00040	Vittoria	Cuzzocrea	1981-11-29	0	+39 3374139081	Canale Battaglia, 231 70100, Bari (BA)	Infermiere	laurea scienze infermieristiche
A00054	Graziella	Chianamonte	1977-02-28	0	+39 3698825289	Stretto Romano, 750 Piano 8 70020, Poggiorsini (BA)	medico	laurea medicina
A00067	Vincenza	Missoni	1983-05-27	0	+39 3759935131	Viale Michelotto, 811 Appartamento 9 70123, San Paolo (BA)	Infermiere	laurea scienze infermieristiche
A00078	Stefania	Majewski	1982-07-04	0	+39 3380878176	Strada Mazzeo, 14 Appartamento 3 70125, Carrassi (BA)	Infermiere	laurea scienze infermieristiche
A00138	Pina	Gibilisco	1986-11-30	0	+39 3686324441	Rotonda Fioravanti, 7 Appartamento 3 70127, Santo Spirito (BA)	Infermiere	laurea scienze infermieristiche
A00166	Tiziana	Casellati	1984-09-26	0	+39 3710487460	Viale Luisa, 538 Piano 1 70121, Bari (BA)	Infermiere	laurea scienze infermieristiche
A00167	Antonina	Ortese	1981-09-09	0	+39 3636485518	Borgo Cerutti, 1 70127, Santo Spirito (BA)	Infermiere	laurea scienze infermieristiche
A00181	Maurilio	Boccherini	1971-04-21	1	+39 3963309466	Incrocio Matteo, 3 Piano 8 70054, Giovinazzo (BA)	caposala	laurea magistrale infermieristica
A00192	Lilla	Opizzi	1985-01-21	0	+39 3951310514	Piazza Ivo, 5 Appartamento 5 70129, Loseto (BA)	Infermiere	laurea scienze infermieristiche
A00235	Graziella	Padovano	1985-11-24	0	+39 3498555316	Strada Adriano, 145 Piano 1 70128, Palese (BA)	Infermiere	laurea scienze infermieristiche
A00272	Fausto	Piacentini	1999-04-21	1	+39 3776491986	Incrocio Mondadori, 6 70056, Molfetta (BA)	medico	laurea medicina
A00293	Pina	Marsili	1985-10-06	0	+39 3506648417	Vicolo Scandone, 774 Piano 2 70056, Molfetta (BA)	medico	laurea medicina
A00309	Cassandra	Moccia	1984-01-26	0	+39 3842982009	Incrocio Marta, 52 Piano 7 70131, Carbonara Di Bari (BA)	Infermiere	laurea scienze infermieristiche
A00337	Stefano	Ravaglioli	1987-01-13	1	+39 3321371565	Borgo Bova, 8 70022, Altamura (BA)	medico	laurea medicina
A00356	Romina	Renault	1985-10-22	0	+39 3348692243	Borgo Forza, 21 70122, Bari (BA)	Infermiere	laurea scienze infermieristiche
A00358	Ludovica	Treccani	1982-08-14	0	+39 3530123293	Contrada Silvia, 30 Appartamento 58 70121, Bari (BA)	Infermiere	laurea scienze infermieristiche
A00386	Nicoletta	Giannini	1983-12-14	0	+39 3223966708	Piazza Adriana, 17 70127, Santo Spirito (BA)	Infermiere	laurea scienze infermieristiche

Se l’utente dovesse selezionare, durante la creazione, una terapia intensiva già creata, allora il sistema chiederà se l’utente vuole cancellare il file contenete il personale già assegnato alla terapia intensiva e crearne una nuova oppure se vuole tornare indietro e selezionare un’altra terapia intensiva.

In caso l’utente decidesse di crearne una nuova, allora il sistema leggerà il file del personale già esistente e, per ciascun dipendente al suo interno, procederà a segnalarlo come non assegnato all’interno del dataset.

Una volta terminato il processo di creazione di una terapia intensiva, l’utente potrà interrogare la KB per visualizzare l’organizzazione del personale assegnato ad una specifica terapia intensiva già creata tramite l’utilizzo della query 10. Se l’utente dovesse provare ad utilizzare la query senza aver creato alcuna terapia intensiva (quindi inserendo come input il nome di una terapia intensiva non ancora creata), la query restituirà come output un messaggio di errore indicando che non è stato trovato alcun risultato relativo alla terapia intensiva indicata.

Strumenti utilizzati

Per l’implementazione della sezione inerente alla creazione e risoluzione dei CSP abbiamo utilizzato la libreria di Python chiamata “python-constraint” [3].

Per ciascuna terapia intensiva sono state utilizzate funzioni fornite dalla libreria per la dichiarazione delle variabili: `csp.addVariable(“nomevariabile”, dominiovariabile);`

per la ricerca del personale per ciascun ruolo durante la compilazione della lista del personale:

`csp.addConstraint(tipovincolo nomivariabili: formulavincolo, (nomivariabilidichiarate));`

per la selezione di un risolutore: `csp.setSolver(nomerisolutore);`

per ottenere l'insieme di soluzioni: `csp.getSolutions()` e
per cancellare i vincoli e variabili utilizzati nella ricerca appena conclusa: `csp.reset()`.

Una volta ottenuta la lista di possibili soluzioni dalla funzione `csp.getSolutions()`, abbiamo deciso di utilizzare un algoritmo esaustivo per la ricerca all'interno del dataset del personale che soddisfi i vincoli relativi alla lista soluzioni ottenuta.

L'efficienza di questa soluzione è discutibile, ma data la sua semplicità e, soprattutto, la certezza che la ricerca è in grado di restituire sempre i risultati attesi, assumendo che essi esistano, abbiamo deciso di implementare questo tipo di algoritmo.

Il suo funzionamento è molto semplice: l'algoritmo legge una riga dalla lista delle soluzioni (la quale rappresenta una possibile soluzione) ottenuta dalla funzione citata precedentemente (precedentemente analizzata affinché la sua formattazione non crei problemi con il funzionamento di questo algoritmo) ed esegue una ricerca all'interno del dataset per trovare un dipendente, ovvero una riga, in cui i valori soddisfano la soluzione letta e il campo "assegnazione" è uguale a zero.

Qualora l'algoritmo dovesse trovare un dipendente che soddisfi queste condizioni, la sua matricola verrà inserita in una lista del personale, la quale verrà successivamente salvata in quanto rappresenterà il personale assegnato ad una specifica terapia intensiva.

A parte il controllo eseguito sul campo assegnazione, e la modifica del suo valore effettuata immediatamente dopo la selezione di un dipendente che soddisfi tutte le condizioni, l'algoritmo ha anche il compito di limitare la ricerca ad un numero preciso di personale, deciso dal numero di posti inserito durante la fase iniziale di creazione della terapia intensiva.

Infine, l'algoritmo deve controllare che, alla scelta di ogni personale, la sommatoria dello stipendio totale del personale selezionato non superi il limite di budget inserito durante la parte iniziale della creazione della terapia intensiva.

Decisioni di Progetto

Il numero di posti letto è stato ipotizzato da noi avere un minimo di 1 ed un massimo di 10.

Le dotazioni dell'organico per ciascuna terapia intensiva sono state definite ipotizzando un determinato numero di personale stimato come necessario all'apertura della terapia intensiva in rapporto al numero dei posti letto che l'utente ritiene di attivare (tra 1 e 3, tra 4 e 7, tra 8 e 10).

Il budget minimo per ciascuna terapia intensiva è stato calcolato moltiplicando la dotazione organica ottenuta come illustrato precedentemente per lo stipendio base (senza considerare anzianità e le extra competenze come gli straordinari e le indennità di turno) ricavato, per ciascuna figura professionale, dai contratti di lavoro vigenti sia per la dirigenza medica che per i paramedici.

Per l'individuazione del personale da assegnare a ciascuna terapia intensiva sono stati scelti i seguenti vincoli.

Per il personale con qualifica di OSS e Infermiere abbiamo deciso di escludere coloro che possiedono limitazioni lavorative in seguito al riconoscimento di handicap psicofisici. Tale scelta è motivata dall'esigenza di dover disporre, nelle unità di terapia intensiva da istituire, considerato il lavoro particolarmente stressante che lì viene svolto, di personale in piena efficienza fisica e che possa svolgere completamente il regolare orario di servizio, senza usufruire delle riduzioni di turno previste dalla legge n.104/1992.

Per il personale infermieristico è stato considerato un ulteriore vincolo relativo alla residenza dettato dalla esigenza, particolarmente sentita nelle terapie intensive, di poter disporre di personale

aggiuntivo in regime di reperibilità. Tale regime consiste nell'obbligo, del personale che, pur essendo a riposo, si rende "pronto disponibile" per tamponare urgenze straordinarie dovute, ad esempio, da un numero superiore e non previsto, di ricoveri in tale terapia. Il suddetto personale deve poter raggiungere il reparto entro 20 min dalla chiamata in reperibilità e, pertanto, è stato posto il vincolo della residenza in una distanza dall'ospedale che possa essere coperta in un tale lasso di tempo. Per questo motivo, sempre ipotizzando la collocazione dell'ospedale in Bari, sono stati individuati i dipendenti che hanno la propria residenza a Bari o nei comuni limitrofi (La lista dei CAP appartenenti ai comuni limitrofi a Bari è composta dai seguenti codici: 70100, da 70121 a 70132). Tale selezione è stata attuata prendendo in riferimento i numeri del codice di avviamento postale (CAP) contenuti negli indirizzi all'interno delle stringhe che compongono il dataset del personale.

Per quanto riguarda il vincolo di esperienza per gli infermieri abbiamo deciso un valore medio di 20 anni; questo valore costituisce l'esatto spartiacque nella carriera lavorativa che ha normalmente la durata di 40 anni. In questa maniera l'utente può decidere se preferire personale infermieristico più giovane (e quindi con maggiore entusiasmo ed energia) oppure più anziano e quindi più esperto. Qualora l'utente abbia optato per un numero di posti letto da 8 a 10, abbiamo preferito includere nell'organico la figura di un caposala che possa coordinare il personale paramedico che, per tale dotazione, è più numeroso, tale da necessitare quindi di un coordinatore. Il vincolo posto per l'individuazione di tale figura è costituito dal titolo; il caposala, per svolgere quella funzione, deve aver conseguito la laurea magistrale in infermieristica e quindi, la scelta di tale figura, potrà accadere solo su chi è già caposala oppure su un infermiere che abbia comunque conseguito il titolo abilitante.

Per essere assegnati alla terapia intensiva da istituire, il vincolo posto ai medici consisterà nel possesso della specializzazione (o equipollente o affine) prevista dalla legge (D.P.R. 483 e 484 già sopra menzionati) selezionata tra quelle maggiormente attinenti all'attività da svolgere all'interno dell'unità intensiva (poste all'interno di una lista, per ciascuna terapia intensiva, denominata "spec_ ***nometerapiaintensiva***").

Per la scelta dei medici responsabili di struttura semplice o complessa, alle medesime condizioni dovrà essere aggiunto il vincolo consistente nell'anzianità di servizio di almeno 5 anni.

In casi particolari alcuni vincoli potrebbero essere trascurati, come ad esempio in dialisi dove, poiché essa è una terapia semi-intensiva non attiva nel turno notturno, non sarà considerato il vincolo sulla residenza per la reperibilità del personale infermieristico.

Un altro caso particolare lo troviamo in Neonatologia, dove non verrà assegnato alcun personale con qualifica di OSS (per il tipo di pazienti ricoverati al suo interno) e sarà invece considerato il vincolo della scelta esclusiva delle sole infermiere donne, in quanto ritenute più idonee alla specificità dell'attività esplicata.

Durante la fase di testing abbiamo deciso di aggiungere a ciascun vincolo, qualora non fosse già presente, la necessità che il titolo qualificate assegnato a ciascun dipendente, il cui valore è presente all'interno del campo titolo, sia uguale al titolo necessario per essere assunto con la qualifica per la quale si sta eseguendo la ricerca.

Ad esempio, se si sta effettuando una ricerca sui medici, a parte il vincolo relativo alla qualifica di medico, è stato aggiunto un vincolo sul valore di titolo, il quale, in questo caso, deve necessariamente essere uguale a laurea medicina.

Questo vincolo non è necessario ai fini della ricerca, ma è stato inserito per rendere più efficiente il funzionamento dell'algoritmo di ricerca, dato che nella lista soluzioni restituita da solver del CSP non vi saranno più tutte le combinazioni che includono valori diversi nel campo titolo.

Per la stessa ragione, è stata utilizzata una metodologia diversa per verificare i vincoli posti sul campo residenza. Da considerare tutti le combinazioni di CAP possibili, abbiamo deciso di codificare

l'appartenenza alla lista di CAP limitrofi a Bari (valutati durante la ricerca di infermieri per le terapie intensive) tramite l'utilizzo di un valore booleano aggiunto in una colonna denominata limitrofi.

Motivazione reset dei vincoli e variabili:

Una volta eseguiti dei test per capire come meglio implementare le funzioni messe a disposizione dalla libreria da noi utilizzata abbiamo notato che la risoluzione di un CSP che includa più vincoli sulle stesse variabili risulta impossibile.

Ad esempio, se in un problema abbiamo due variabili A (con dominio $[1,2,3]$) e B (con dominio $[4,5,6]$) e vengono applicati due vincoli del tipo: 1. $(A = 1)$ e 2. $(A = 2 \text{ and } B = 5)$, allora la lista delle soluzioni sarà nulla, in quanto la condizione tale da soddisfare il vincolo 1 sarà in contraddizione con la condizione necessaria per soddisfare il vincolo 2 (A non può essere sia uguale ad 1 che uguale a 2). Perciò abbiamo deciso, per ciascuna figura professionale, di dichiarare i vincoli relativi ad essa, eseguire la ricerca ed eliminare, una volta ottenuto il risultato, i vincoli e le variabili precedentemente dichiarati.

Questo ci permetterà di eseguire, per ciascun CSP relativo ad una terapia intensiva, una serie di ricerche per ogni ruolo al suo interno, dichiarando ed eliminando i vincoli e variabili relativi a ciascuna ricerca per evitare conflitti.

Purtroppo, data la struttura della libreria utilizzata e le metodologie fornite da essa, non è stato possibile effettuare una singola dichiarazione delle variabili relative al problema, aggiungere vincoli relativi alle variabili, effettuare la ricerca, rimuovere i vincoli utilizzati per la ricerca e aggiungerne nuovi. Per risolvere questo problema, abbiamo deciso di cancellare tutte le variabili e vincoli relativi alla ricerca terminata e dichiarare nuovamente le variabili e vincoli per la ricerca successiva. Anche se questa soluzione è molto semplice e poco elegante, non abbiamo notato alcun problema con le performance e la qualità delle soluzioni ottenute.

Risultati e Performance

Come accennato precedentemente, l'algoritmo di ricerca da noi utilizzato ci assicura, poiché esaustivo, che, se nel dataset è presente un dipendente i cui valori relativi alle proprie caratteristiche soddisfano le condizioni associate alla riga della lista criteri presa in considerazione durante l'iterata, allora la matrice del dipendente verrà inserita all'interno della lista personale relativa alla terapia intensiva che si sta istituendo.

Data la natura dell'algoritmo, però, possiamo constatare che le sue performance non sono ottimali. L'algoritmo andrà infatti a iterare la lettura delle 500 righe che compongono il dataset per ciascuna riga della lista criteri ottenuta dal solver del CSP associato alla creazione della terapia intensiva indicata all'inizio del processo.

Durante la nostra sperimentazione dell'algoritmo abbiamo osservato che la sua durata varia principalmente a causa della quantità di variabili considerate durante la dichiarazione del CSP associato. Questo significa che per alcune ricerche eseguite su liste criteri ottenute ponendo dei vincoli abbastanza specifici, su più variabili, ovvero le caratteristiche di ciascun dipendente all'interno del dataset, l'algoritmo necessiterà di iterare più volte la lettura completa del dataset fino a trovare una corrispondenza tra le condizioni presenti in lista criteri ed i valori associati ad un dipendente all'interno del dataset.

Questo processo di selezione del personale verrà ripetuta per ciascuna qualifica all'interno della terapia intensiva per il numero relativo del personale necessario, calcolato a partire del numero di posti letto inseriti all'inizio del processo; questo significa che l'algoritmo non terminerà una volta trovato un solo dipendente che soddisfi le condizioni, ma sarà necessario trovarne la quantità specificata dalla variabile ed essa associata.

Durante la nostra fase di test, il processo di creazione di una terapia intensiva ha avuto una durata che varia dai 2 ai 3 minuti. Qualora fossero stati inseriti valori non sufficienti per terminare la creazione della terapia intensiva, il programma, una volta terminate le iterate sulla lista criteri ottenuta dal CSP associato, visualizzerà un messaggio di errore invitando l'utente ad aumentare il budget per terminare la creazione della terapia intensiva selezionata.

Apprendimento Non-Supervisionato

Sommario

La tecnica di clustering è stata scelta per osservare come un algoritmo di apprendimento non-supervisionato riesca a raggruppare il personale dell'ospedale in base alle caratteristiche, in particolare qualifica, anzianità e classificazione, le quali incidono notevolmente sul valore dello stipendio associato a ciascun dipendente.

L'obiettivo di questa tecnica è l'individuazione di cluster, ovvero classi costituite da esempi simili in base al calcolo di valori particolari denominati centroidi.

Per assegnare il valore del cluster calcolato in relazione al dipendente esaminato, abbiamo deciso di aggiungere una colonna nel dataset denominata "cluster", all'interno della quale verranno memorizzati i valori durante la fase di apprendimento.

Il clustering può essere di due tipi: Hard clustering (assegnazione di ciascun esempio ad una singola classe di appartenenza) e Soft clustering (distribuzione di appartenenza di tutti gli esempi rispetto a ciascuna classe).

Per questo progetto abbiamo deciso di implementare un metodo generale per l'Hard clustering, ovvero il K-Means.

Una volta eseguito il modulo del clustering, l'utente potrà interrogare la base di conoscenza tramite la query 9; questa query, se eseguita prima del processo di clustering, restituirà un messaggio di errore, indicando all'utente di eseguire prima il processo di apprendimento non-supervisionato all'interno del modulo apposito.

Strumenti utilizzati

Per l'implementazione del K-Means abbiamo deciso di utilizzare la libreria scikit-learn [4].

Per codificare le caratteristiche categoriche in caratteristiche numeriche, affinché il K-Means possa funzionare correttamente, abbiamo deciso di utilizzare la tecnica del One-Hot Encoding [5], la quale creerà una variabile binaria univoca per ciascuna categoria presente all'interno del dataset.

Decisioni di Progetto

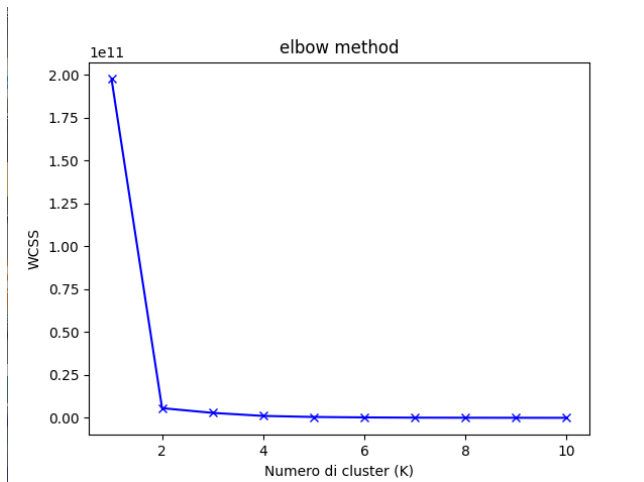
Prima di eseguire la fase di codifica, abbiamo deciso di creare un file denominato "**datasetForClustering.csv**", il quale verrà poi utilizzato durante il processo di clustering.

Questa versione del dataset contiene, per ogni dipendente, solo tre delle feature presenti all'interno del dataset completo, ovvero quelle che noi abbiamo ritenuto fossero più influenti sull'aspetto remunerativo del personale; le feature scelte sono state: Qualifica, Anzianità e Classificazione.

Terminata la fase di codifica delle variabili di natura categorica, lo step successivo è consistito nella determinazione del numero ottimale di cluster, ovvero il numero di gruppi con il quale verranno raggruppati tutti i dipendenti.

Per questo step abbiamo deciso di utilizzare la tecnica del gomito (Elbow-Method) [6].

In questa tecnica verranno rappresentati, all'interno di un grafico, sull'asse delle ordinate i valori delle somme dei quadrati intra-cluster (WCSS), mentre sull'asse delle ascisse verranno rappresentati i k cluster.



Osservando il grafico, abbiamo deciso di considerare 2 come valore dei k cluster.

Le principali metriche scelte per la configurazione del modello K-Means sono le seguenti:

N_Cluster = 2, ovvero il numero dei cluster in cui il dataset verrà raggruppato;

Random_state = 42, questo parametro permette di controllare la riproducibilità dei risultati, in particolare, più alto è il valore di questo parametro, più è probabile che il modello fornisca gli stessi risultati qualora venga addestrato con gli stessi dati.

N_init = 10, questo parametro indica il numero di volte che verrà eseguito l'algoritmo con diversi valori iniziali relativi ai centroidi.

L'algoritmo selezionerà la soluzione migliore tra le esecuzioni effettuate sulla base della somma dei quadrati delle distanze. Più alto è il valore, più alte saranno le probabilità di ottenere una soluzione di qualità migliore (a costo di un maggior tempo di calcolo).

Il risultato del clustering è stato inserito in una versione aggiornata del dataset denominata **"dataset+Cluster.csv"**, la quale ha una colonna aggiuntiva rispetto al dataset originale denominata Cluster, al cui interno verrà inserito il numero del cluster in cui è stato raggruppato ciascun dipendente.

Valutazione

Per la valutazione delle prestazioni del modello abbiamo utilizzato due metriche:

WCSS (Within-Cluster Sum of Squares): questa metrica, come accennato in precedenza, calcola la somma dei quadrati delle distanze di ogni esempio rispetto al centroide del cluster assegnato ad esso assegnato. Un valore più basso di WCSS indica una maggiore coesione all'interno dei cluster.

Silhouette Score: questa metrica valuta, come la WCSS, la coesione all'interno dei cluster. Essa è calcolata per ciascun esempio e rappresenta il rapporto tra la distanza media al cluster ad esso assegnato e la distanza media ai cluster più vicini. Uno score più alto indica una migliore separazione dei cluster.

I risultati da noi ottenuti, con un valore di k cluster = 2, sono i seguenti:

```
WCSS: 5688374128.912514
Silhouette Score: 0.9352837909412306
```

Questi risultati ci mostrano che vi è un'ottima separazione tra i due cluster (notato dal valore elevato del Silhouette Score), ma, osservando il valore troppo alto del WCSS, non vi è una divisione abbastanza omogenea dei dati all'interno dei raggruppamenti.

Conclusioni

Questo progetto ha avuto come obiettivo la realizzazione, a partire da un dataset da noi creato, di un sistema intelligente che rappresenti il dominio d'interesse da noi selezionato.

La conoscenza di fondo sulla quale è stata creata la Knowledge Base è costituita dal dataset, al cui interno sono presenti 500 dipendenti di un ipotetico ospedale, creato appositamente, con l'aiuto di un esperto del dominio, per essere il più possibile coerente con l'organizzazione di un ospedale reale con le stesse caratteristiche.

Successivamente, abbiamo creato KB utilizzando la logica del primo ordine, in particolare tramite l'utilizzo di Prolog. Abbiamo quindi definito regole e query per interrogare il sistema e ottenere informazioni specifiche riguardanti le qualifiche, gli stipendi e le informazioni anagrafiche dei dipendenti.

Una volta costruita la KB, abbiamo deciso di implementare un modulo che, tramite la risoluzione di un CSP, permetta all'utente di creare una terapia intensiva, rappresentata da una lista del personale contenente i dipendenti che l'algoritmo ha deciso di assegnare alla specifica terapia intensiva.

Infine, per studiare in maniera più mirata come i valori relativi al salario totale di ciascun dipendente possa essere utilizzato per classificare tutto il personale del dataset, abbiamo utilizzato la tecnica K-Means di clustering, la quale raggruppa tutti i dipendenti all'interno del dataset in un numero k specifico di gruppi (cluster).

Dai risultati ottenuti possiamo osservare che esistono principalmente due gruppi di dipendenti, legato principalmente alla differenza di salario tra il personale para-medico ed il personale medico, il quale possiede uno stipendio totale maggiore (quindi considerando anche gli incrementi di anzianità e i bonus relativi alla classificazione medica).

Durante lo sviluppo di questo progetto, ci sono state due funzionalità che non siamo riusciti ad includere. La prima è l'implementazione di un modulo del sistema che utilizzi un'ontologia per rispondere, fornendo informazioni aggiuntive, ad interrogazioni dell'utente riguardanti unità operative, terapie intensive, qualifiche, titoli e, più in generale, altre caratteristiche presenti all'interno della KB.

La seconda funzionalità che non è stata inserita, soprattutto per questioni di tempo, ma è stata sviluppata in parte, è la possibilità di modificare la lista restituita dal modulo di creazione di terapie intensive. Questa funzionalità permetterebbe all'utente di aprire la lista del personale associata ad una delle terapie intensive create e decidere se rimuovere o sostituire alcuni dipendenti al suo interno tramite una ricerca all'interno del dataset che sfrutti la lista di criteri restituita dal CSP solver durante la creazione della stessa terapia intensiva che l'utente sta modificando.

Riferimenti Bibliografici

- [1] <https://faker.readthedocs.io/en/master/>
- [2] <https://pypi.org/project/pyswip/>
- [3] <https://python-constraint.github.io/python-constraint/>
- [4] <https://scikit-learn.org/stable/modules/clustering.html#k-means>
- [5] <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder>
- [6] <https://stackoverflow.com/questions/41540751/sklearn-kmeans-equivalent-of-elbow-method>