

A First Experiment with an Intelligent System for Therapy Recommendation for Osteoporotic Patients

Maxim Nitsenko

Università degli Studi di Ferrara

Supervisor
Prof. **Guido Scavicco**

17 Dicembre 2020



Università
degli Studi
di Ferrara

Si presenta il mio lavoro di tirocinio in cui abbiamo sviluppato un sistema intelligente per la raccomandazione medica per il centro di menopausa e osteoporosi di Ferrara.
Il sistema estende l'esistente software per la gestione dei referti prodotti dal centro, consigliando il tipo di terapia e principio attivo.

Si parte con una panoramica dei sistemi di supporto alla decisione, successivamente parlerò dell'algoritmo impiegato per l'estrazione di regole, poi si tratta del preprocessamento dati, in seguito approfondisco dell'architettura generale del sistema, dopo di che si parlerà delle regole e infine risultati statistici e conclusioni generali.

Nella mia ricerca ho individuato tre categorie principali di CDSS con diverse sottocategorie.

- Knowledge Based CDSS

Tassonomia dei sistemi di supporto alla decisione

Nella mia ricerca ho individuato tre categorie principali di CDSS con diverse sottocategorie.

- Knowledge Based CDSS
 - Regole

Nella mia ricerca ho individuato tre categorie principali di CDSS con diverse sottocategorie.

- Knowledge Based CDSS
 - Regole
 - Ontologie

Nella mia ricerca ho individuato tre categorie principali di CDSS con diverse sottocategorie.

- Knowledge Based CDSS

- Regole
- Ontologie
- Case base

Nella mia ricerca ho individuato tre categorie principali di CDSS con diverse sottocategorie.

- Knowledge Based CDSS
 - Regole
 - Ontologie
 - Case base
- Non knowledge based CDSS

Nella mia ricerca ho individuato tre categorie principali di CDSS con diverse sottocategorie.

- Knowledge Based CDSS
 - Regole
 - Ontologie
 - Case base
- Non knowledge based CDSS
 - Non interpretabile

Nella mia ricerca ho individuato tre categorie principali di CDSS con diverse sottocategorie.

- Knowledge Based CDSS
 - Regole
 - Ontologie
 - Case base
- Non knowledge based CDSS
 - Non interpretabile
 - Regole

Tassonomia dei sistemi di supporto alla decisione

Nella mia ricerca ho individuato tre categorie principali di CDSS con diverse sottocategorie.

- Knowledge Based CDSS
 - Regole
 - Ontologie
 - Case base
- Non knowledge based CDSS
 - Non interpretabile
 - Regole
- Sistemi ibridi

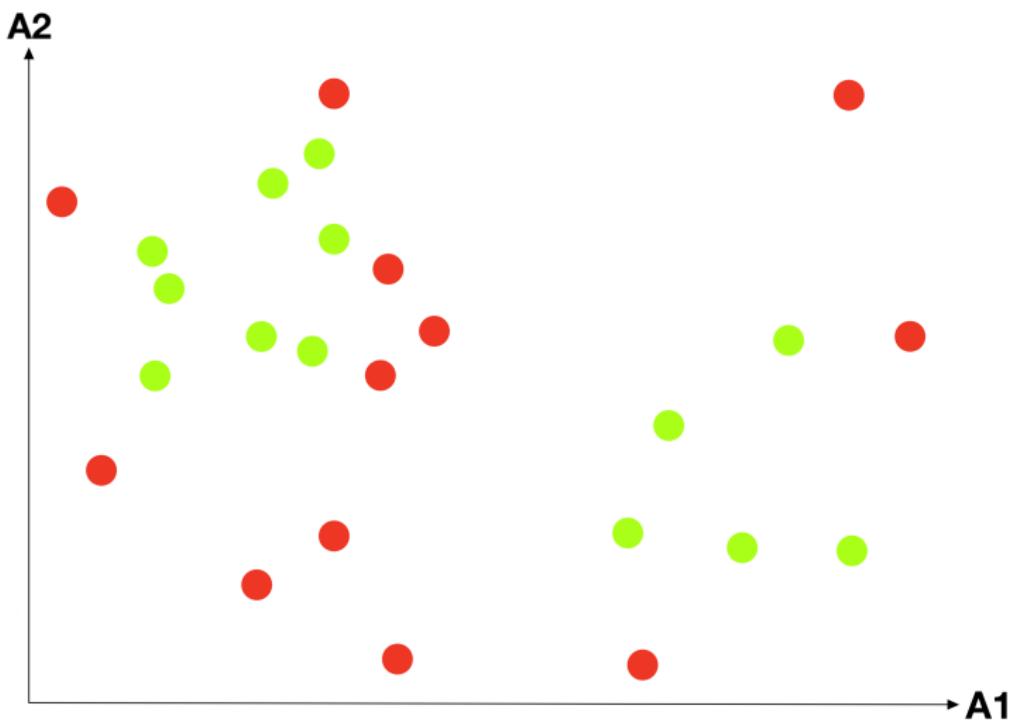
Il nostro scopo era ricavare automaticamente le regole per prevedere la raccomandazione terapeutica.

In più l'interpretabilità dei risultati è fondamentale per l'accettazione prodotto dallo staff medico.

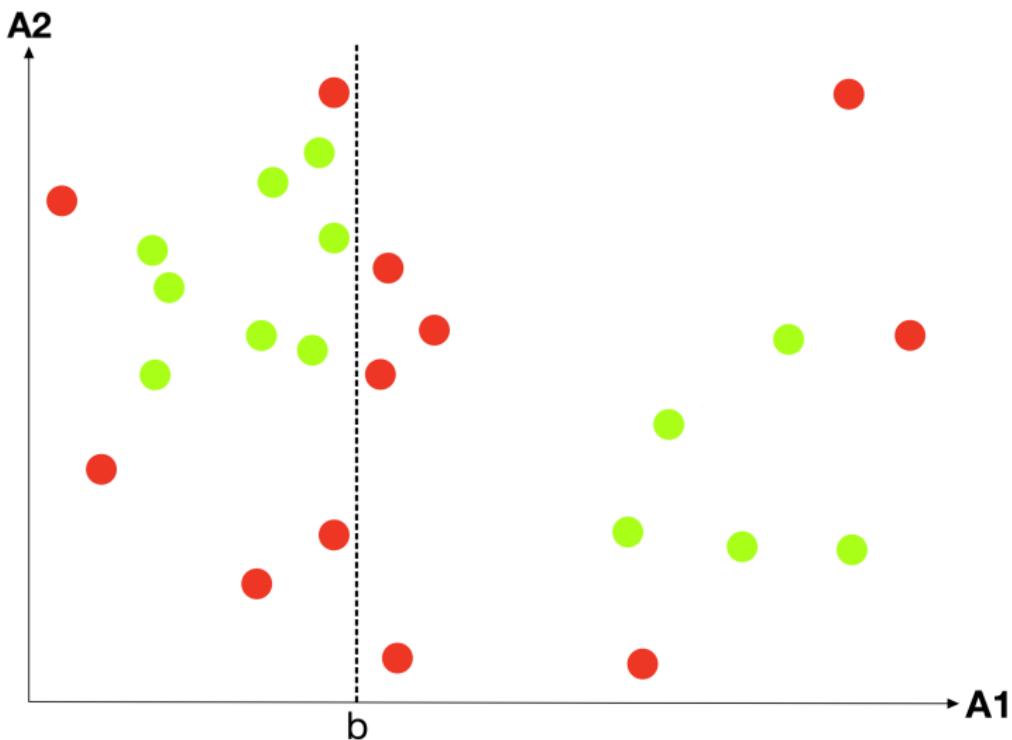
Pertanto la soluzione impiegata è JRip, un estrattore automatico di regole interpretabili dai dati.

L'idea generale di JRip si basa sul sequential covering. Si individua una buona regola secondo un'euristica e si rimuovono tutti gli esempi coperti da essa, sia positivi che negativi. Si itera questo procedimento fino ad esaurire tutti gli esempi positivi.

Funzionamento JRip

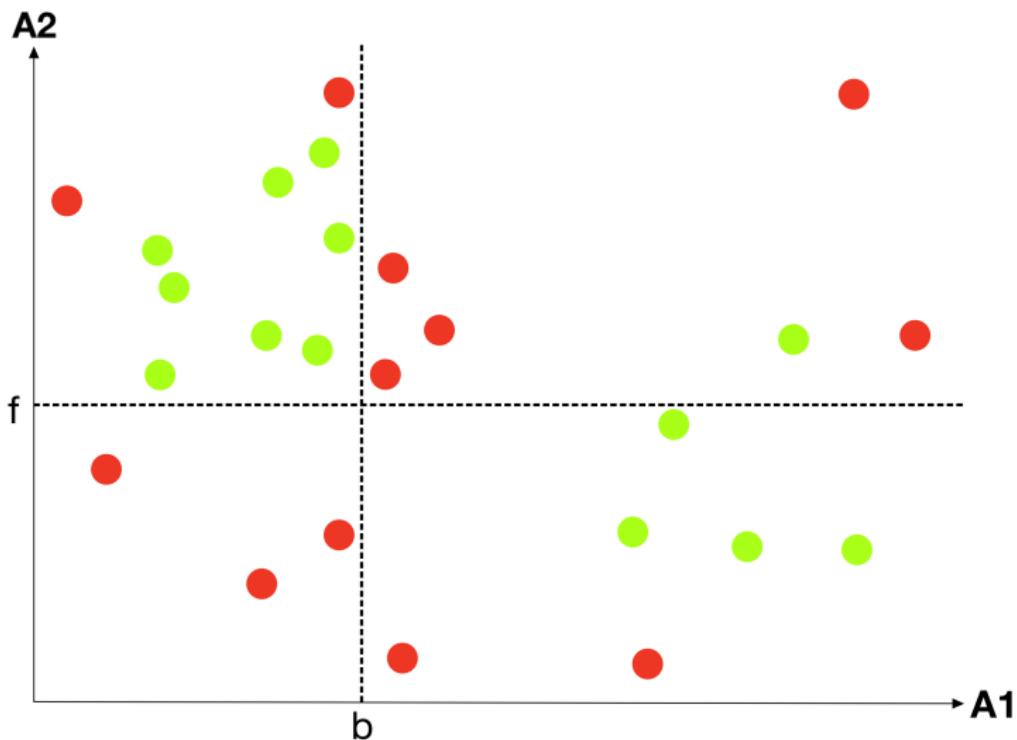


Funzionamento JRip



$$r_1 : (A1 < b)$$

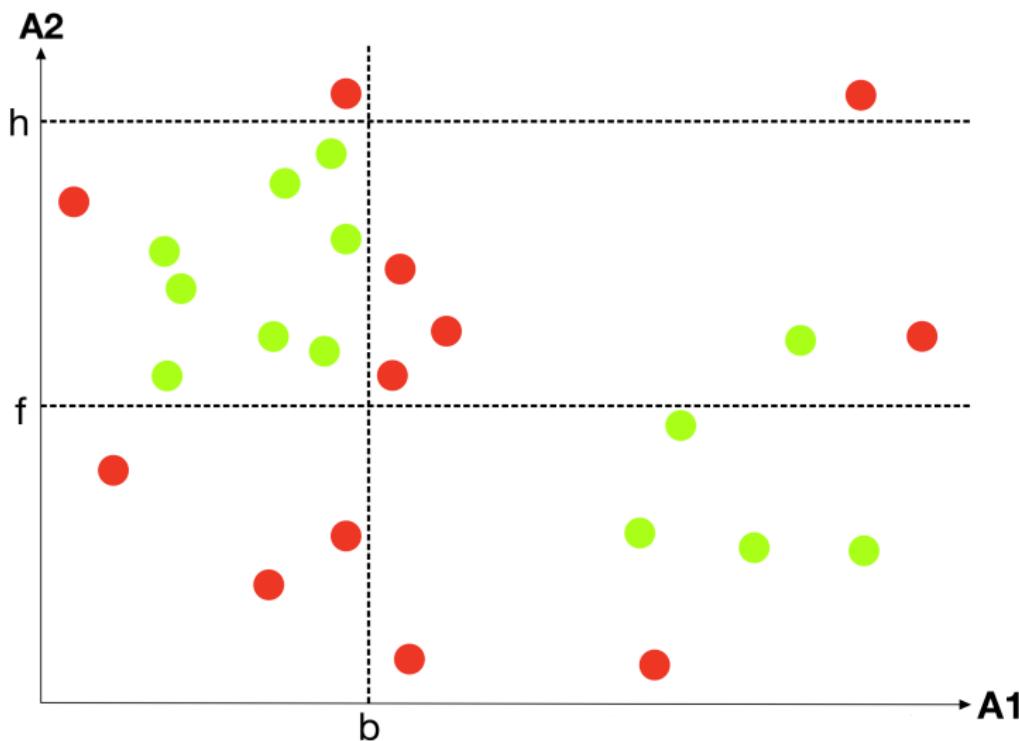
Funzionamento JRip



$$r_1 : (A1 < b \wedge A2 > f)$$



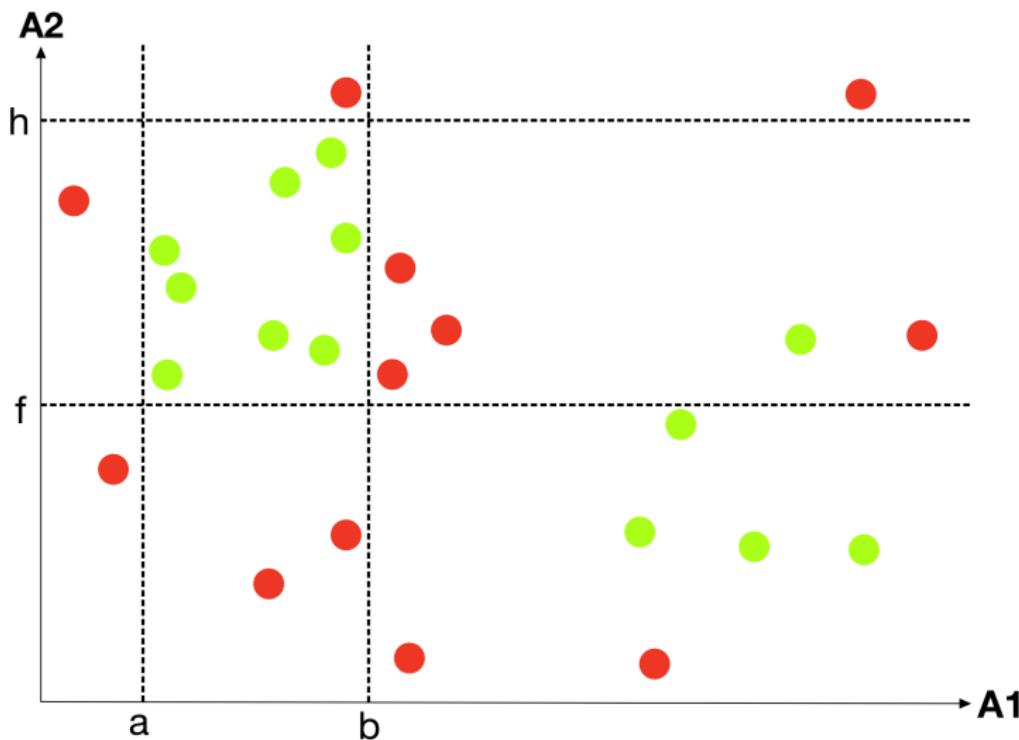
Funzionamento JRip



$$r_1 : (A1 < b \wedge A2 > f \wedge A2 < h)$$



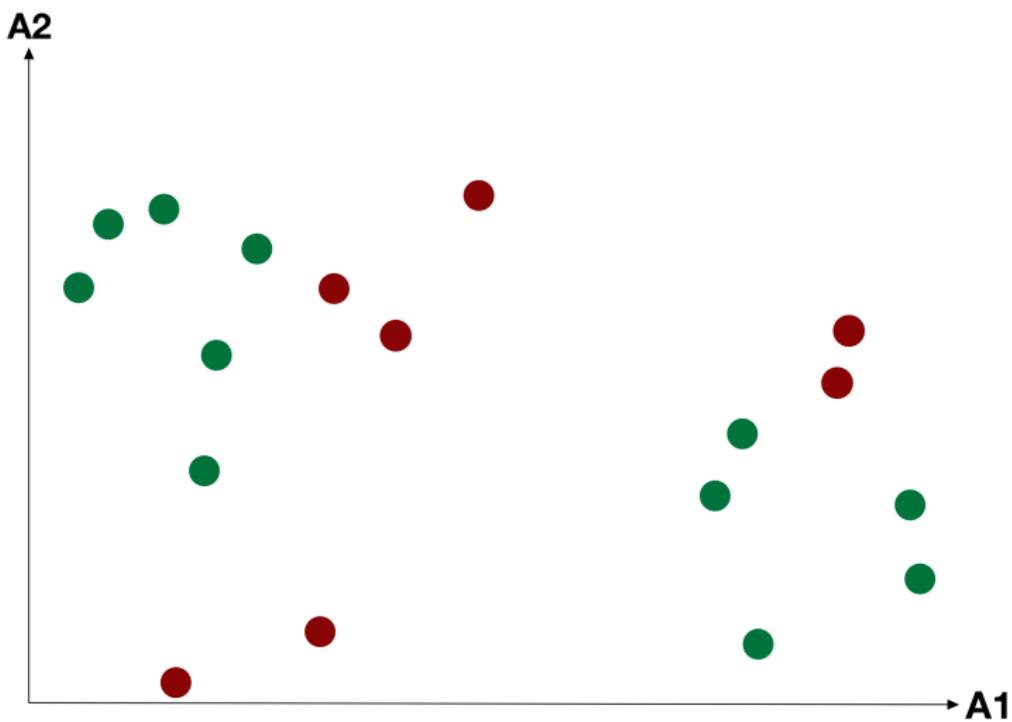
Funzionamento JRip



$r_1 : (A1 < b \wedge A2 > f \wedge A2 < h \wedge A1 > a) \rightarrow \text{●}$



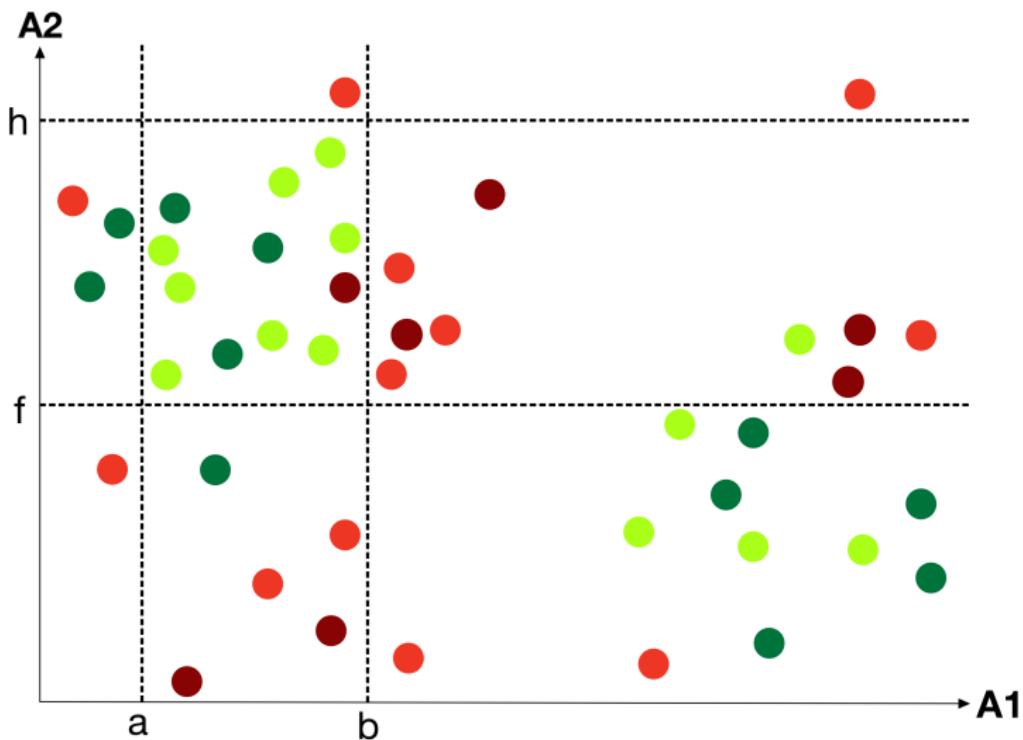
Funzionamento JRip



http://www.csie.ntu.edu.tw/~cjlin/libSVM/



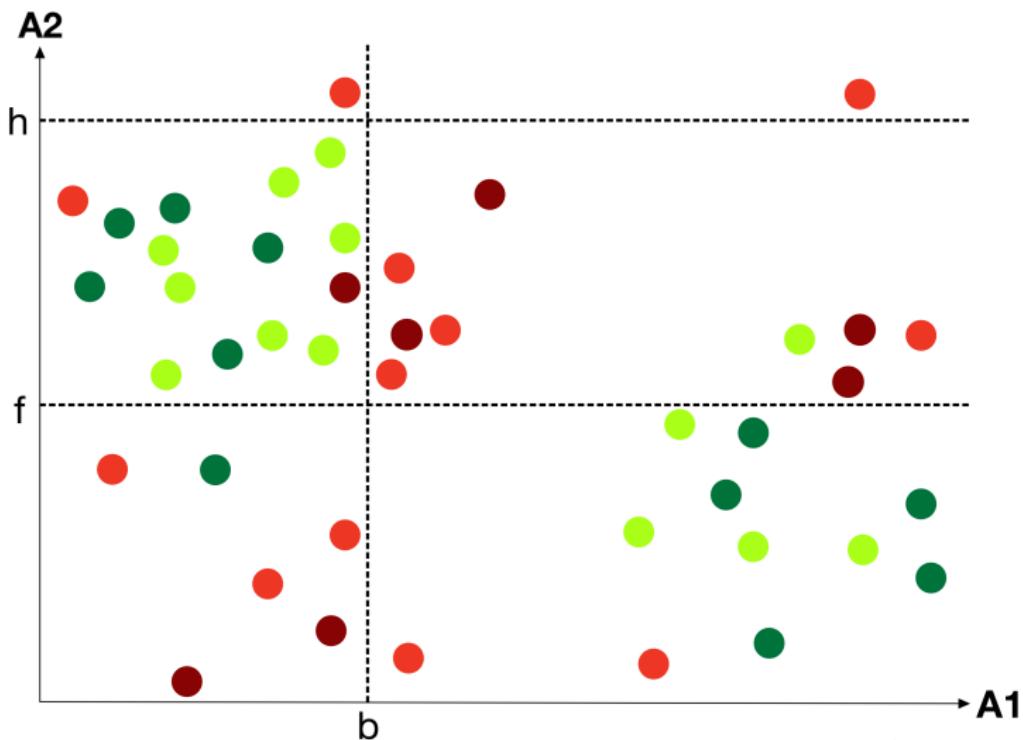
Funzionamento JRip



$r_1 : (A1 < b \wedge A2 > f \wedge A2 < h \wedge A1 > a) \rightarrow \bullet$



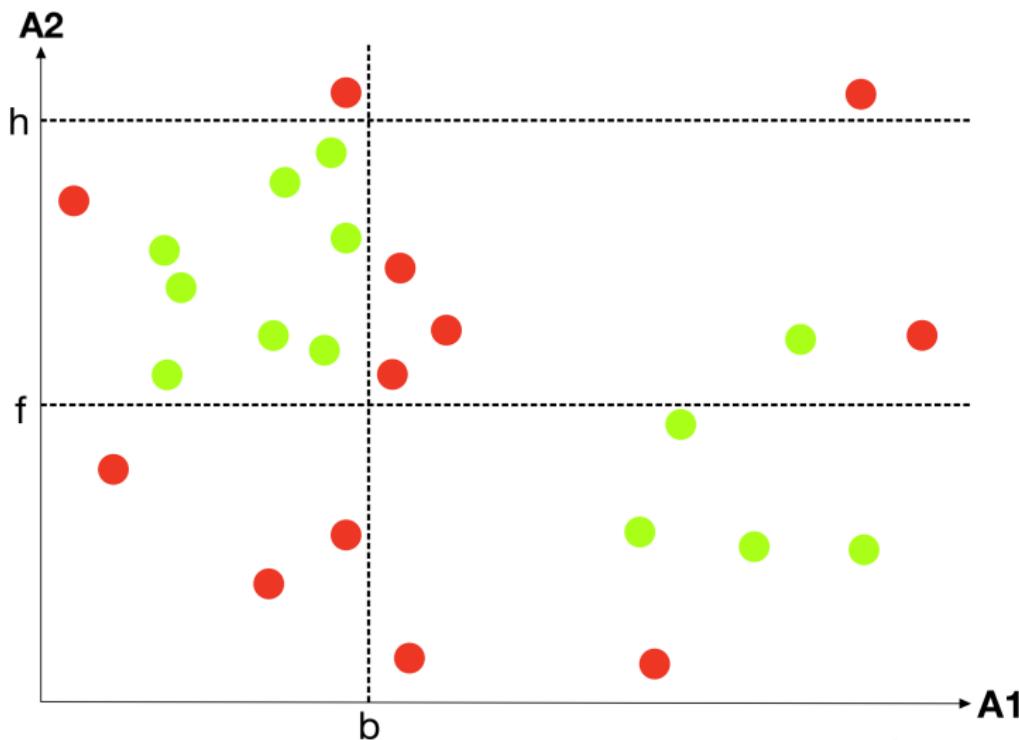
Funzionamento JRip



$$r_1 : (A1 < b \wedge A2 > f \wedge A2 < h) \rightarrow \text{light green circle}$$



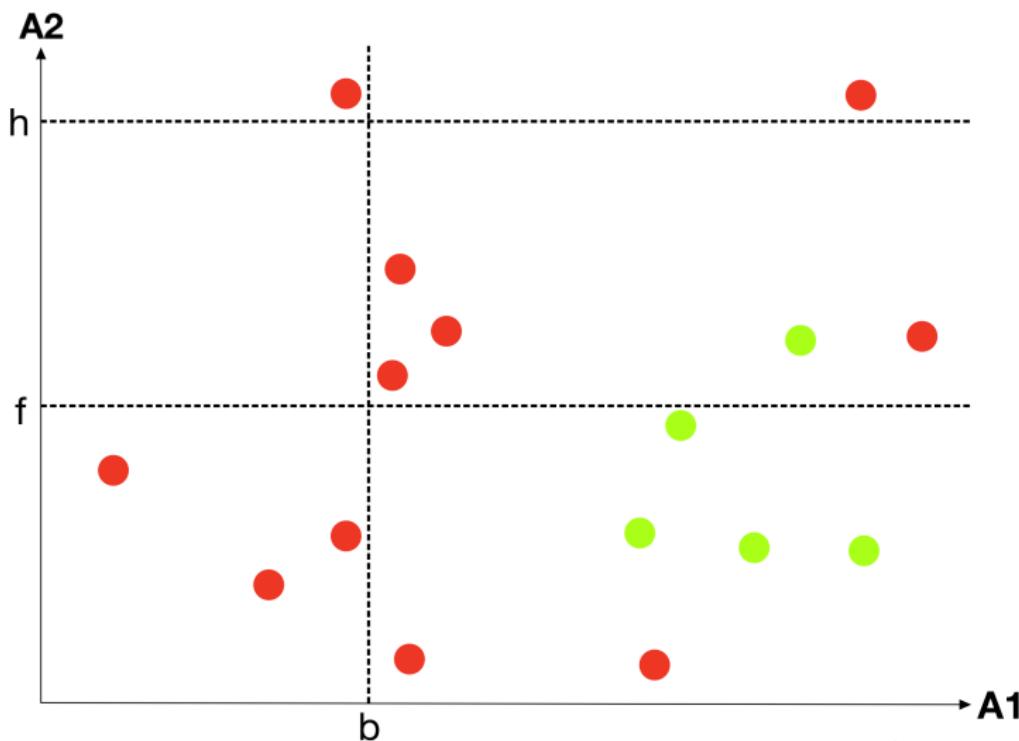
Funzionamento JRip



$r_1 : (A1 < b \wedge A2 > f \wedge A2 < h) \rightarrow \bullet$



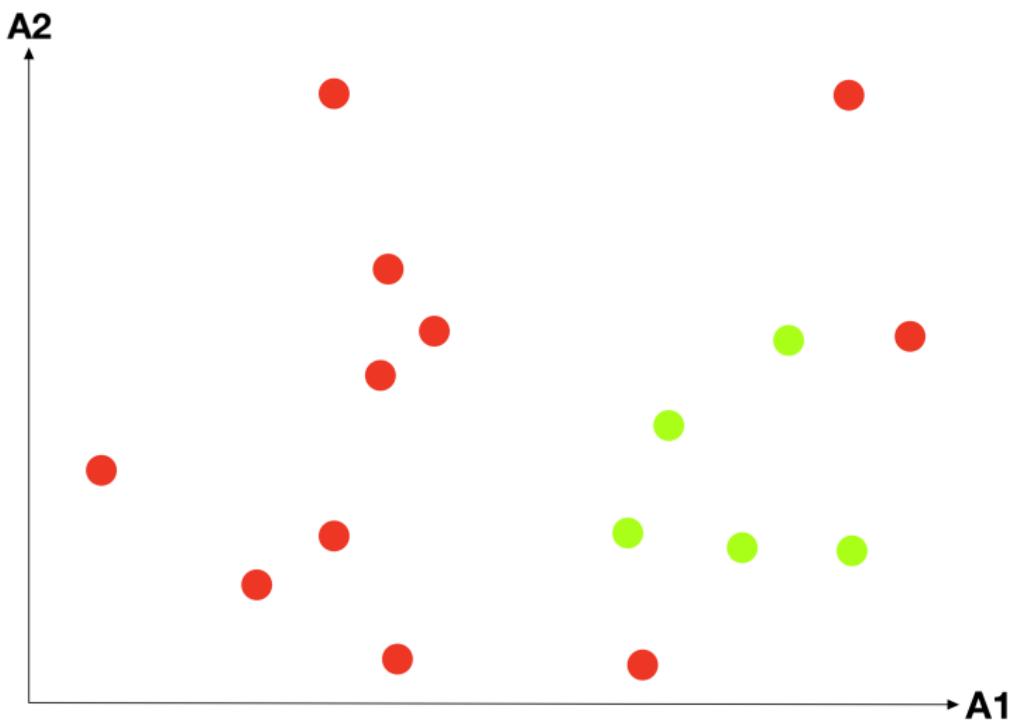
Funzionamento JRip



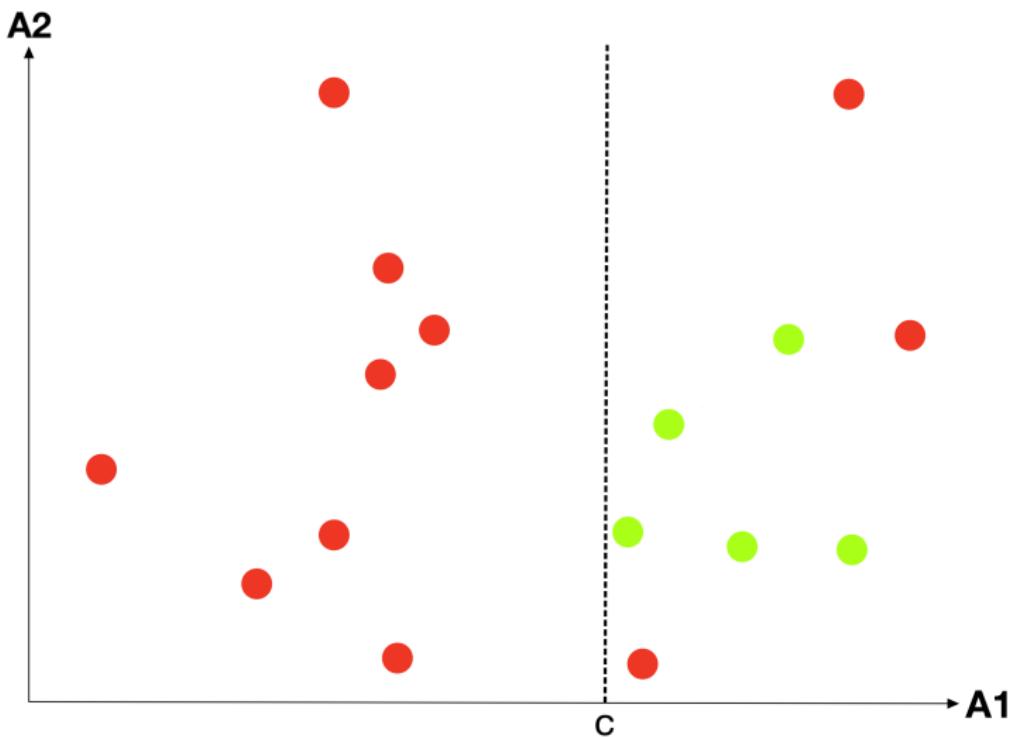
$$r_1 : (A1 < b \wedge A2 > f \wedge A2 < h) \rightarrow \text{Green}$$



Funzionamento JRip

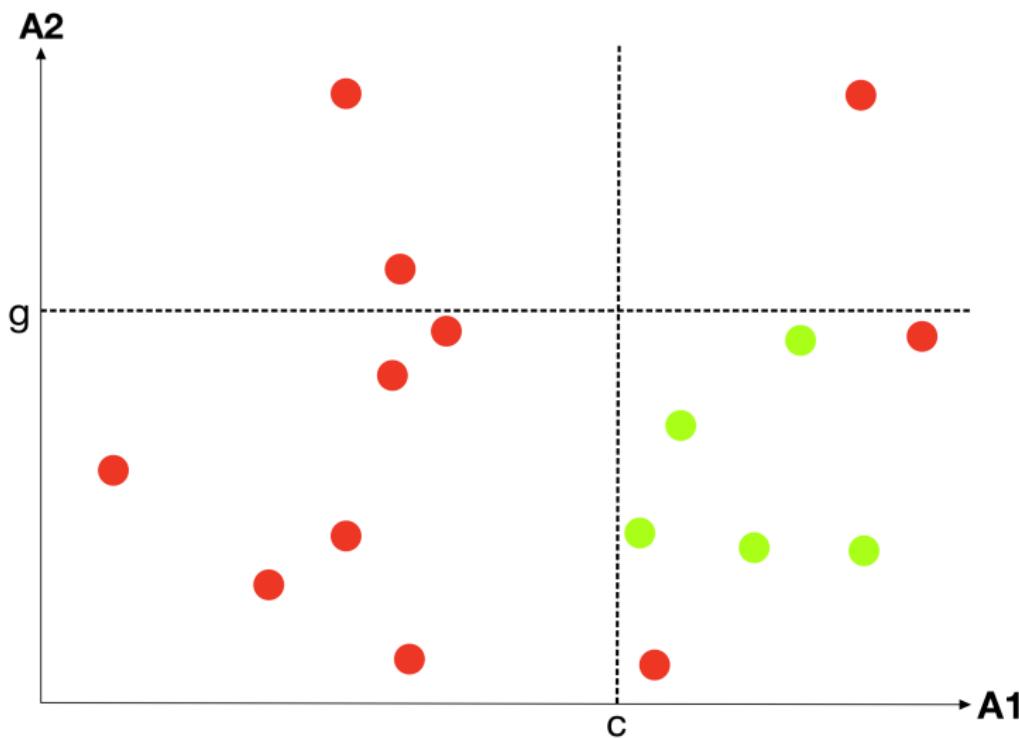


Funzionamento JRip



$$r_2 : (A1 > c$$

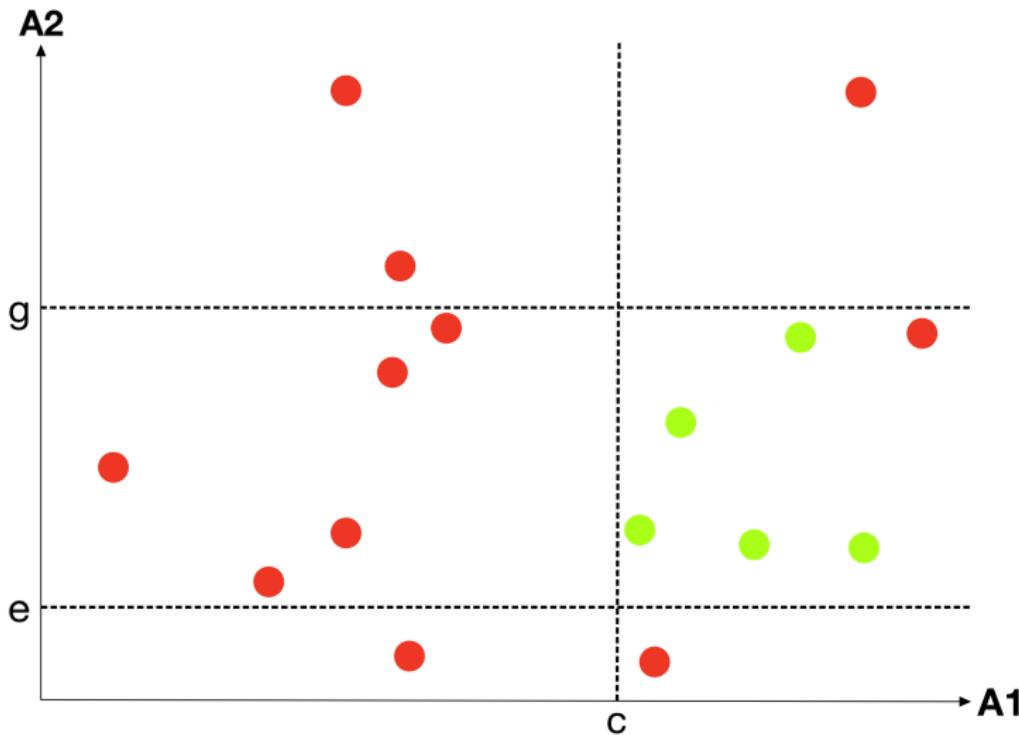
Funzionamento JRip



$$r_2 : (A1 > c \wedge A2 < g)$$



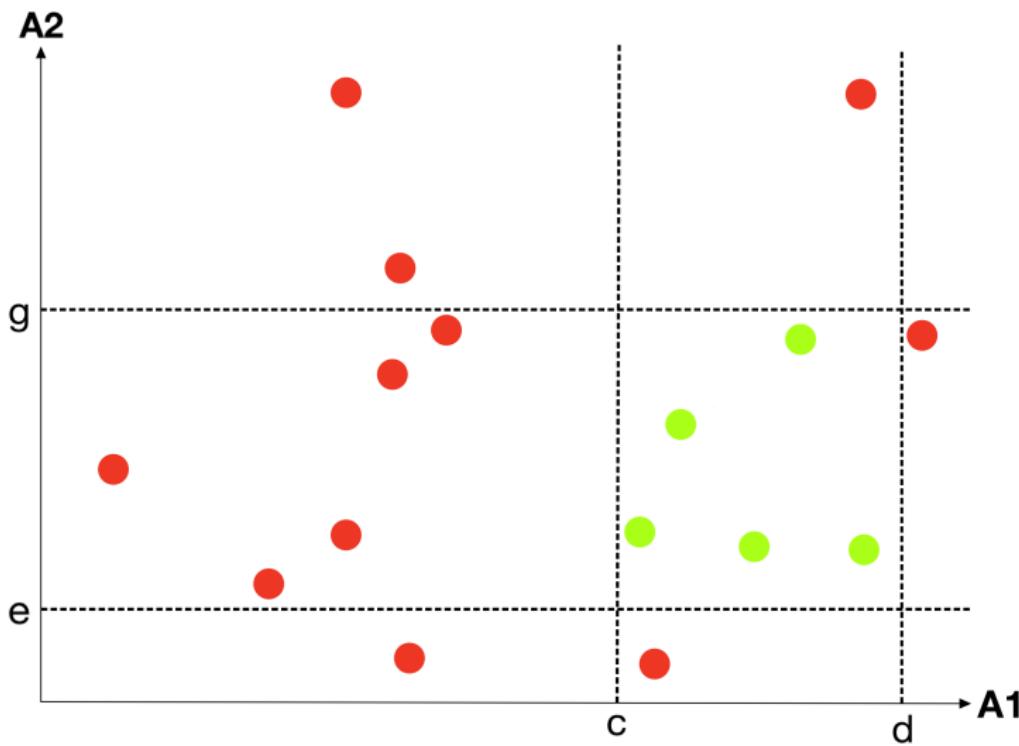
Funzionamento JRip



$$r_2 : (A1 > c \wedge A2 < g \wedge A2 > e)$$



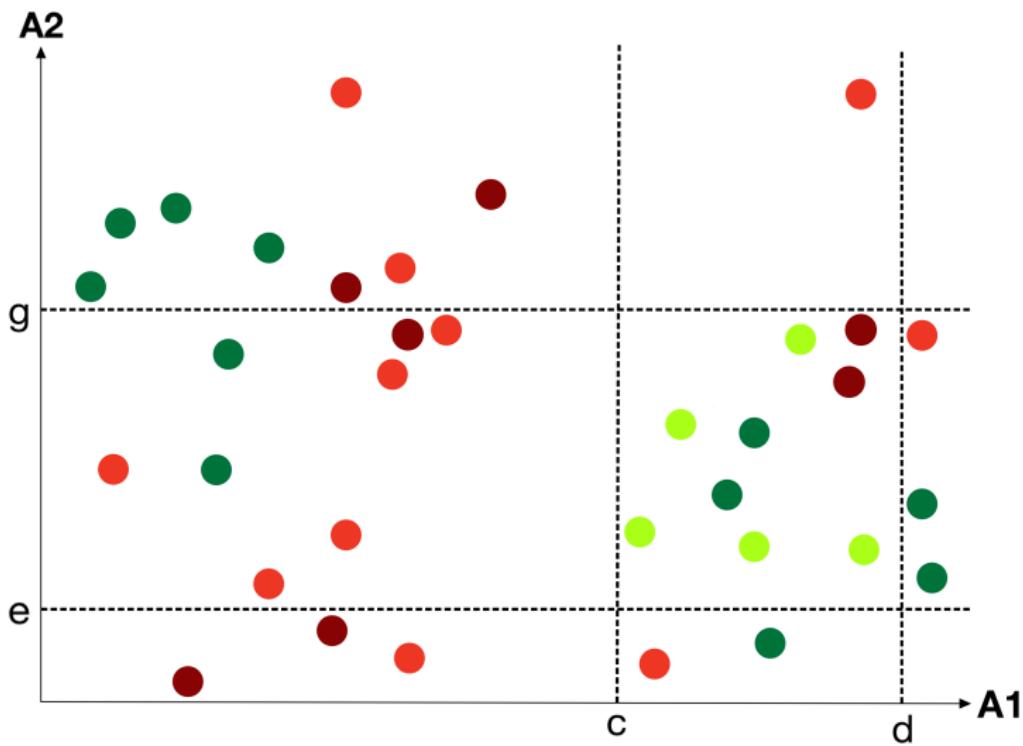
Funzionamento JRip



$r_2 : (A_1 > c \wedge A_2 < g \wedge A_2 > e \wedge A_1 < d) \rightarrow \text{●}$



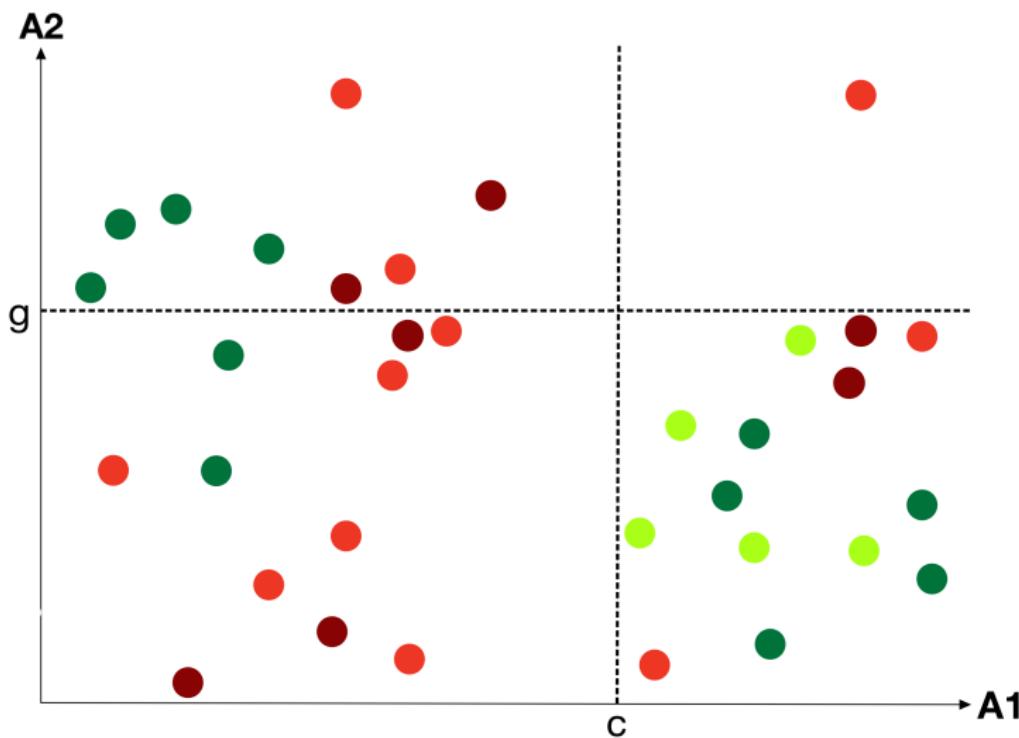
Funzionamento JRip



$$r_2 : (A_1 > c \wedge A_2 < g \wedge A_2 > e \wedge A_1 < d) \rightarrow \text{Yellow Circle}$$



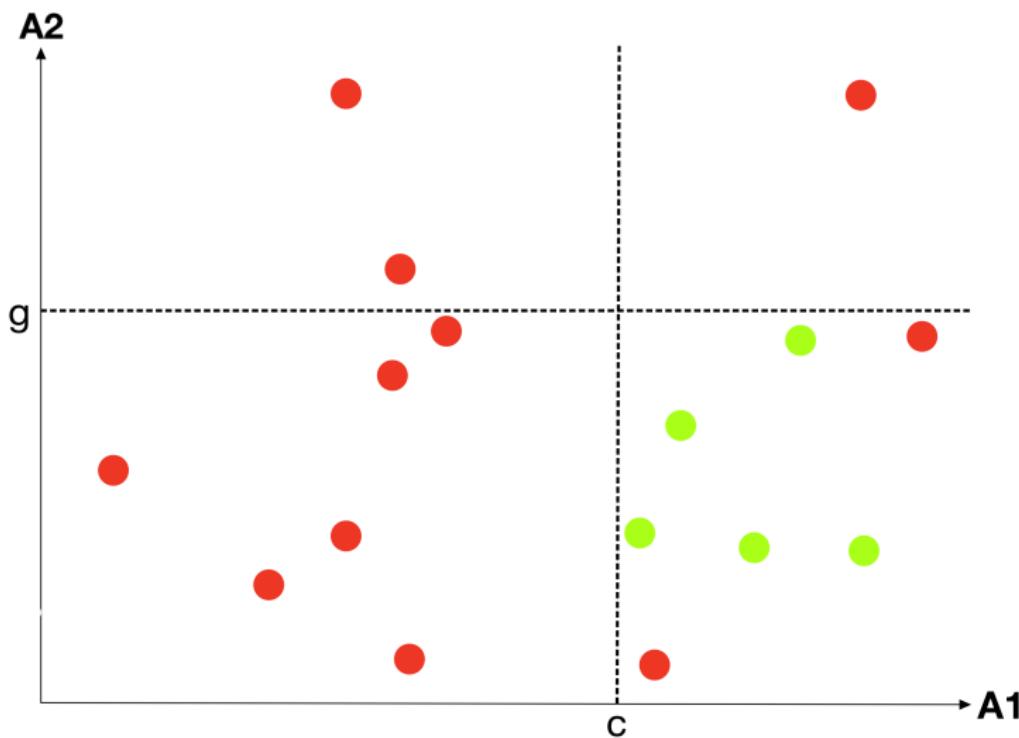
Funzionamento JRip



$$r_2 : (A1 > c \wedge A2 < g) \rightarrow \text{Yellow-Green}$$



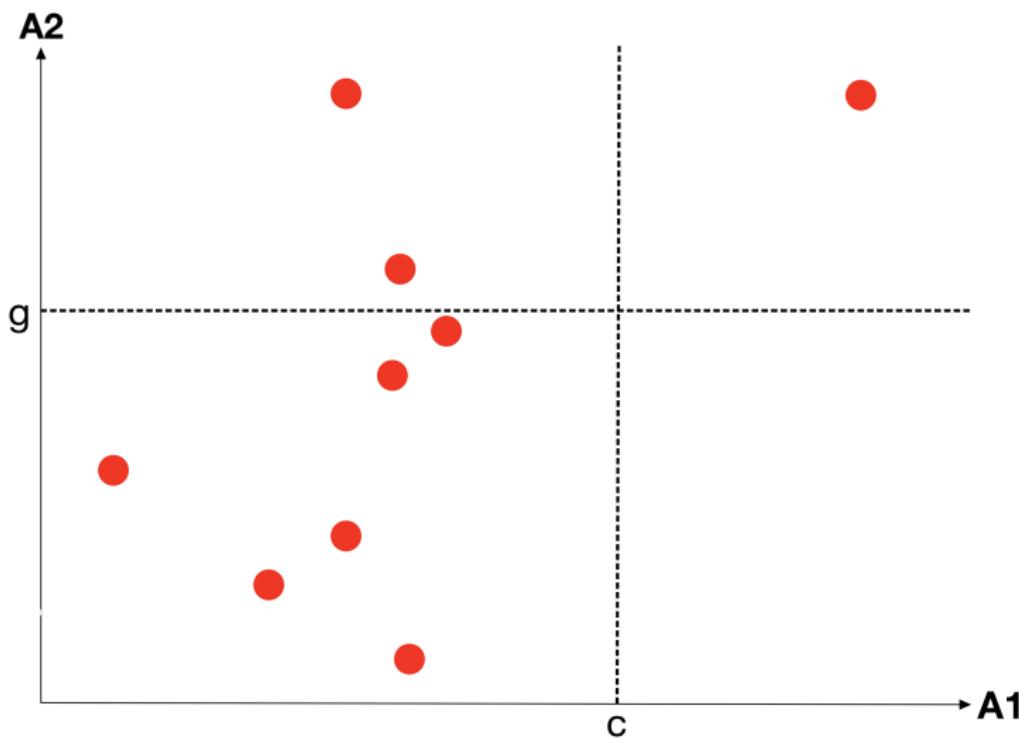
Funzionamento JRip



$$r_2 : (A1 > c \wedge A2 < g) \rightarrow \text{Green Circle}$$

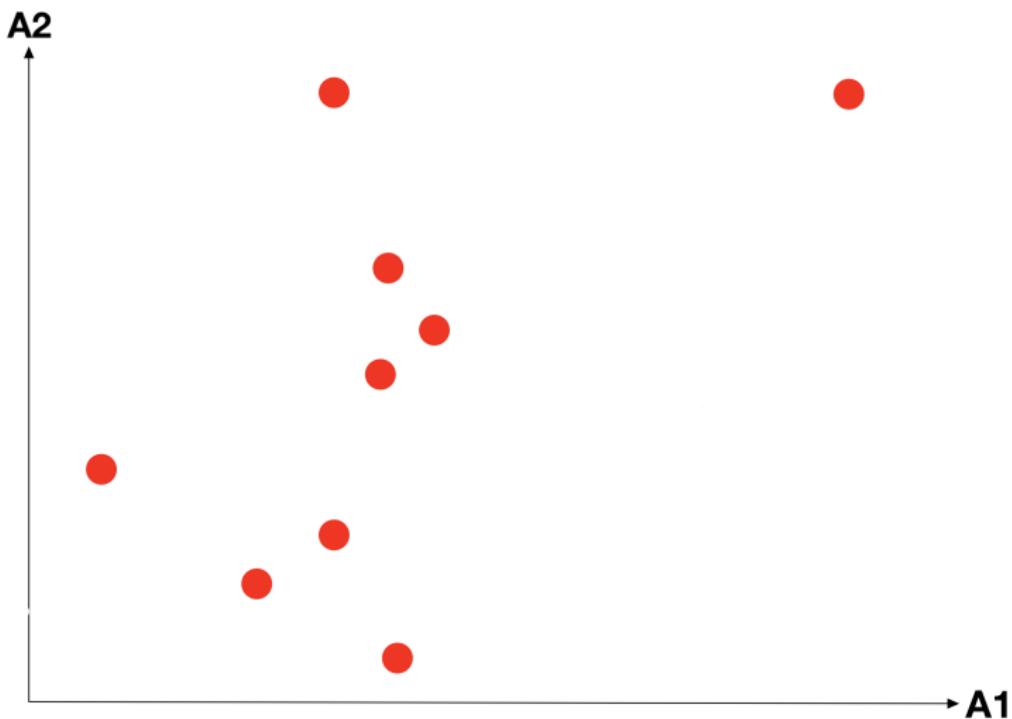


Funzionamento JRip



$$r_2 : (A1 > c \wedge A2 < g) \rightarrow \text{green}$$

JFunzionamento Rip



$r_3 : () \rightarrow \bullet$



Funzionamento JRip

- $r_1 : (A1 < b \wedge A2 > f \wedge A2 < h \wedge A1 > a) \rightarrow \bullet$
- $r_2 : (A1 > c \wedge A2 < g \wedge A2 > e \wedge A1 < d) \rightarrow \bullet$
- $r_3 : () \rightarrow \bullet$

Fase indispensabile in cui si preparano e organizzano i dati prima di avviare l'algoritmo di apprendimento.

La qualità del modello è in relazione diretta con la qualità dei dati.

- Rimozione attributi superflui

Fase indispensabile in cui si preparano e organizzano i dati prima di avviare l'algoritmo di apprendimento.

La qualità del modello è in relazione diretta con la qualità dei dati.

- Rimozione attributi superflui
- Nuovi attributi

Fase indispensabile in cui si preparano e organizzano i dati prima di avviare l'algoritmo di apprendimento.

La qualità del modello è in relazione diretta con la qualità dei dati.

- Rimozione attributi superflui
- Nuovi attributi
- Pulizia dati

Fase indispensabile in cui si preparano e organizzano i dati prima di avviare l'algoritmo di apprendimento.

La qualità del modello è in relazione diretta con la qualità dei dati.

- Rimozione attributi superflui
- Nuovi attributi
- Pulizia dati
- Altro

One-hot-encoding

Si crea un attributo per ogni elemento del dominio della colonna originale. Il nuovo attributo conterrà 1 se l'elemento associato alla colonna compare nella riga corrispondente, altrimenti 0.

The diagram illustrates the process of one-hot encoding. On the left, a vertical table lists categories: 'Intolleranze' (ASA, Noci, Crostacei, Bifosfonati), 'Noci' (ASA, Noci, Crostacei, Bifosfonati), and 'Bifosfonati' (ASA, Noci, Crostacei, Bifosfonati). An arrow points from this list to a larger table on the right. The right table has four columns: 'ASA', 'Noci', 'Crostacei', and 'Bifosfonati'. It contains binary values (0 or 1) indicating the presence of each category in the corresponding row. For example, the first row (ASA) has a 1 in the 'ASA' column and 0s in the others. The second row (Noci) has 0s in the first three columns and a 1 in the 'Bifosfonati' column.

Intolleranze	ASA	Noci	Crostacei	Bifosfonati
ASA	1	0	0	0
Noci, Crostacei	0	1	1	0
Bifosfonati	0	0	0	1
Noci	0	1	0	0

N-grammi

Un n-gramma è una sottosequenza di n elementi di una data sequenza.

Risulta più efficace di singole parole perchè si tiene conto del contesto.

"frattura costale traumatica ottobre 2018"

uni-gramma	frattura	costale	traumatica	ottobre	2018
bi-gramma	frattura costale	costale traumatica	traumatica ottobre	ottobre 2018	
tri-gramma	frattura costale traumatica		costale traumatica ottobre	traumatica ottobre 2018	

Pre-processamento testo

Per produrre risultati più accurati è comune applicare un pre-processamento specifico sul segmento testuale.

- Normalizzazione

Pre-processamento testo

Per produrre risultati più accurati è comune applicare un pre-processamento specifico sul segmento testuale.

- Normalizzazione
- Rimozione stopword

Pre-processamento testo

Per produrre risultati più accurati è comune applicare un pre-processamento specifico sul segmento testuale.

- Normalizzazione
- Rimozione stopword
- Stemming

Pre-processamento testo

Per produrre risultati più accurati è comune applicare un pre-processamento specifico sul segmento testuale.

- Normalizzazione
- Rimozione stopword
- Stemming
- Lowercasing

N-grammi + one hot encoding

Esempio completo della vettorizzazione del campo ALTRO.

ALTRO	ALTRO	attual	terapia	eutirox	tiroide	ibsa	rischi	frattur	signif	attual terapia	terapia eutirox	...
ATTUALMENTE IN TP CON EUTIROX	attual terapia eutirox	ALTRO_1	ALTRO_2	ALTRO_3	ALTRO_4	ALTRO_5	ALTRO_6	ALTRO_7	ALTRO_8	ALTRO_9	ALTRO_10	...
TERAPIA CON TIROIDE IBSA	attual terapia tiroide ibsa	1	1	1	0	0	0	0	0	1	1	...
Rischio di frattura attuale significativo	rischi frattur attual signific	1	1	0	1	1	0	0	0	1	0	...
		1	0	0	0	0	1	1	1	0	0	...

Questa soluzione complica notevolmente il progetto. Ripiego a ngrammi costruiti manualmente.

comuni_PATOLOGIE_UTERINE_DIAGNOSI = ['isterectomia',
'fibromi', 'cisti', 'endometriosi']

comuni_ALTRE_PATOLOGIE = ['vit d', 'eutirox']

:

Interfaccia web sistema intelligente

Terapie ormonali

Suggerisci

Tipo Tipo

aggiungi **cancella**

Terapie osteoprotettive

Suggerisci

Tipo Tipo

aggiungi

Tipo

alendronato 70 mg, 1cpr/settimana

risedonato 75 mg, 1cpr die per due giorni consecutivi al mese

ibandronato 150 mg, 1cpr al mese

clodronato fiale 200 mg im,1 fl ogni 15 giorni

alendronato 70 mg, cpr affervescenti, 1 cpr alla settimana

Alendronato +Vit D in associazione precostituita 1 cpr 70 mg+560 /settimana

Raloxifene, cpr 60 mg, 1 cpr/di in continua

Bazedoxifene cpr 20 mg, 1 cpr/di in continua

Denosumab 1 fl sotto cute, ogni 6 mesi

Teriparatide 20 mcg/die sotto cute

clodronato fiale 200 mg, 1 fl ofni 7 giorni

Zoledronato 5mg e.v, 1 somministrazione annuale

risedronato 35 mg, 1 cpr alla settimana

Vitamina D Terapia

Suggerisci

Tipo Tipo

aggiungi

Vitamina D Supplementazione

Suggerisci

Tipo Tipo

aggiungi **cancella**

Calcio supplementazione

Suggerisci

Tipo Tipo

aggiungi **cancella**

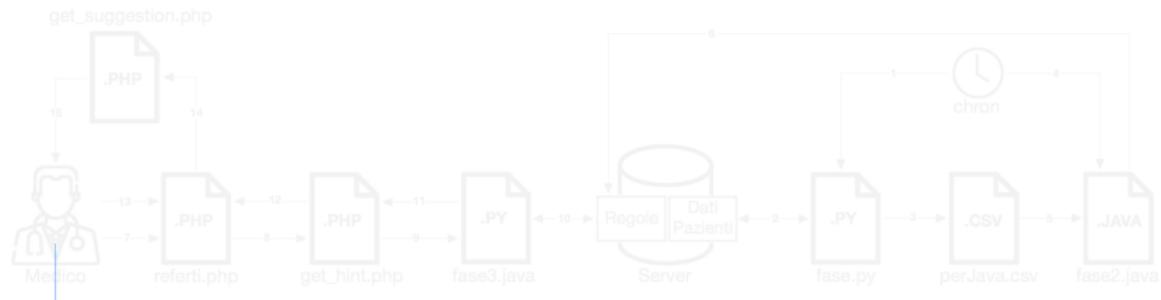


Numero classificatori

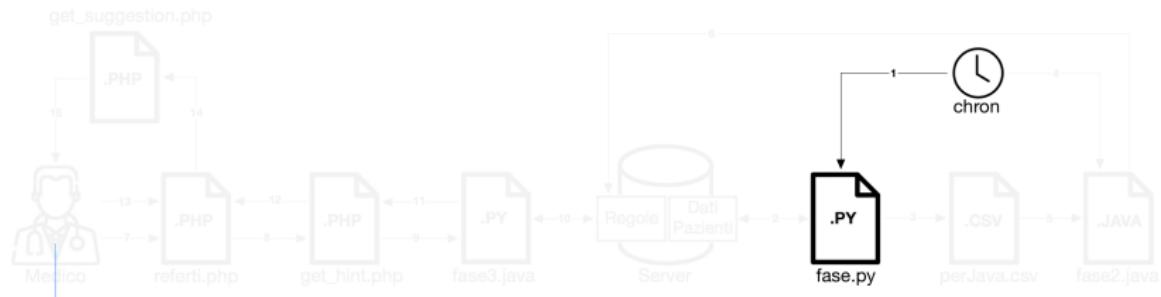
5 classificatori per terapia +
5 classificatori per principio attivo

10 classificatori distinti

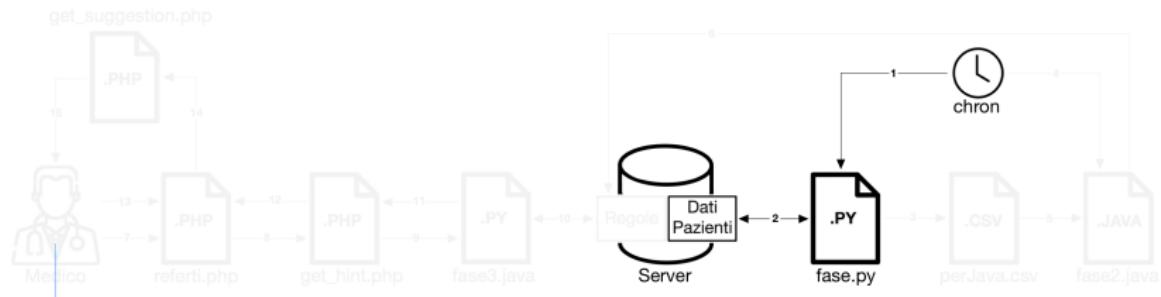
Ciclo?



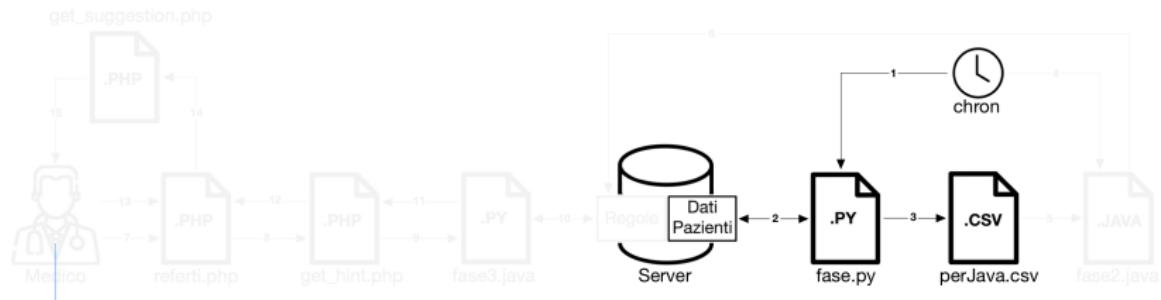
Ciclo?



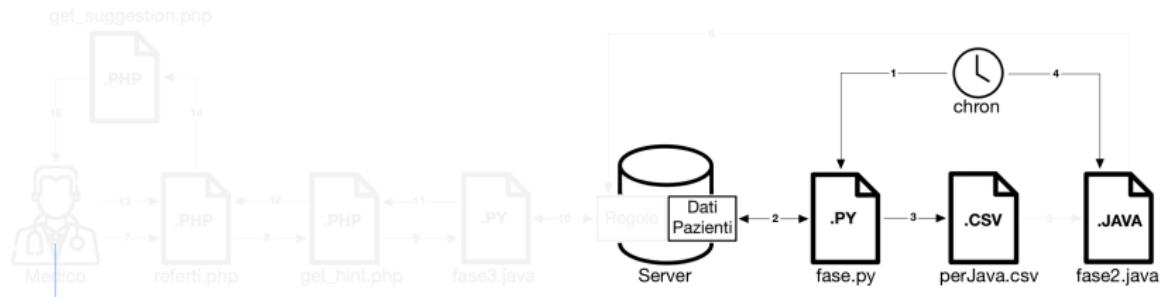
Ciclo?



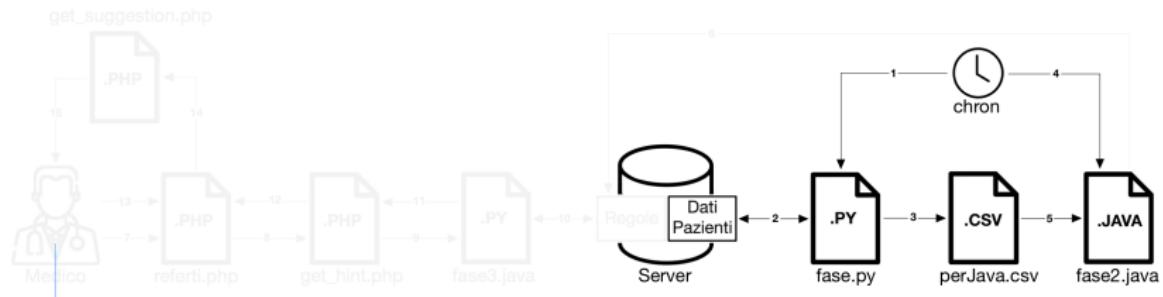
Ciclo?



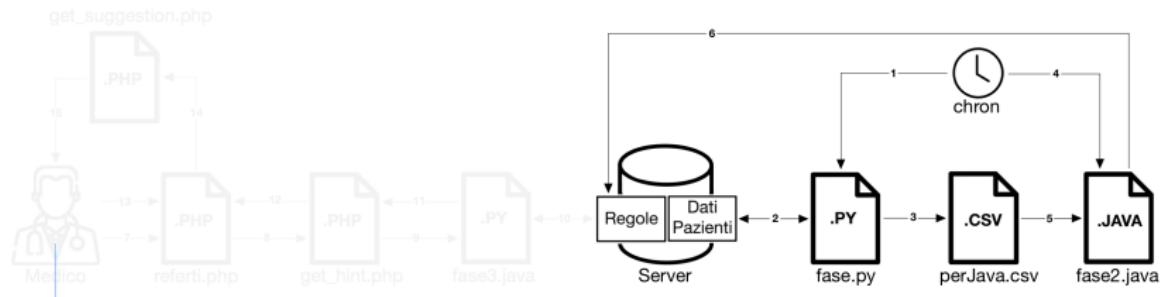
Ciclo?



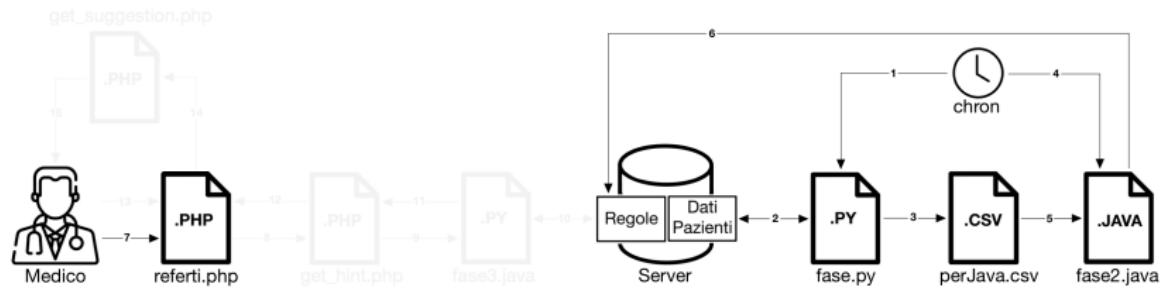
Ciclo?



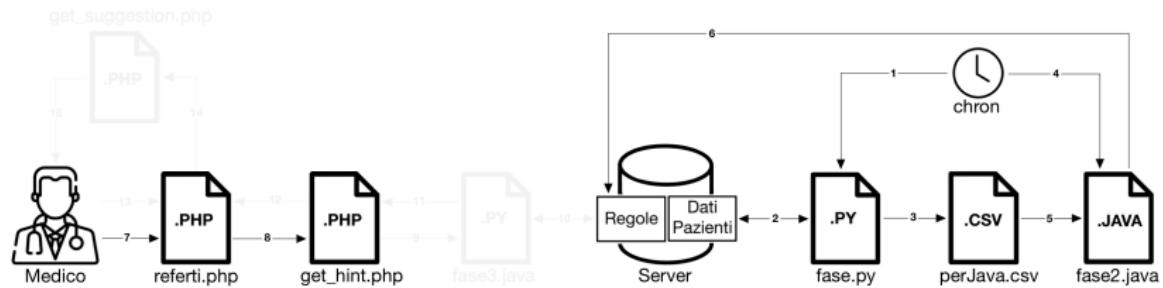
Ciclo?



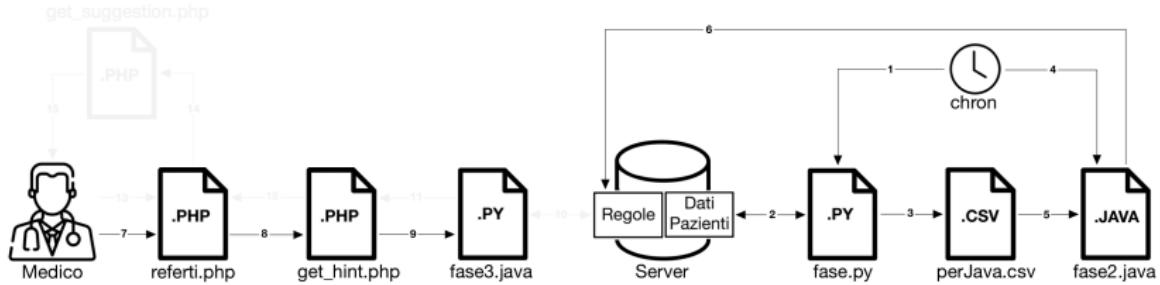
Ciclo?



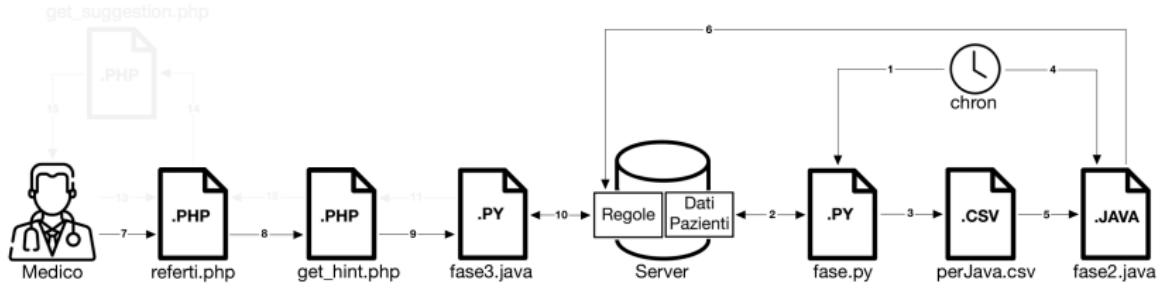
Ciclo?



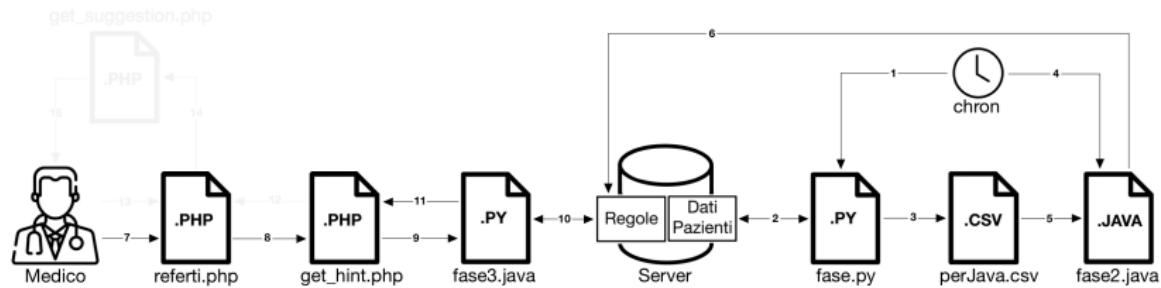
Ciclo?



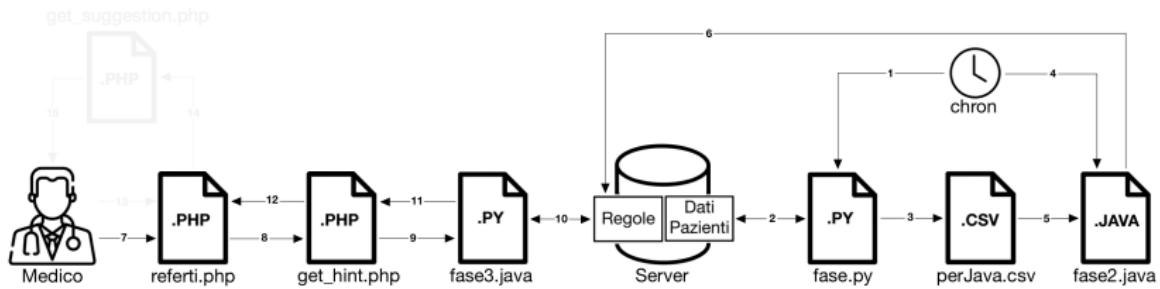
Ciclo?



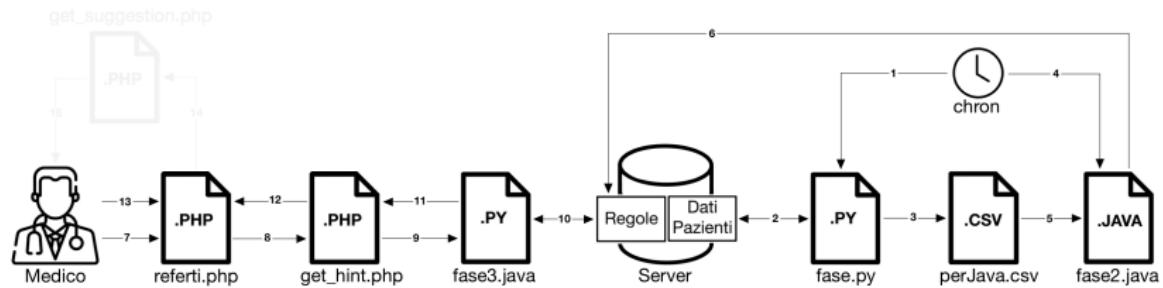
Ciclo?



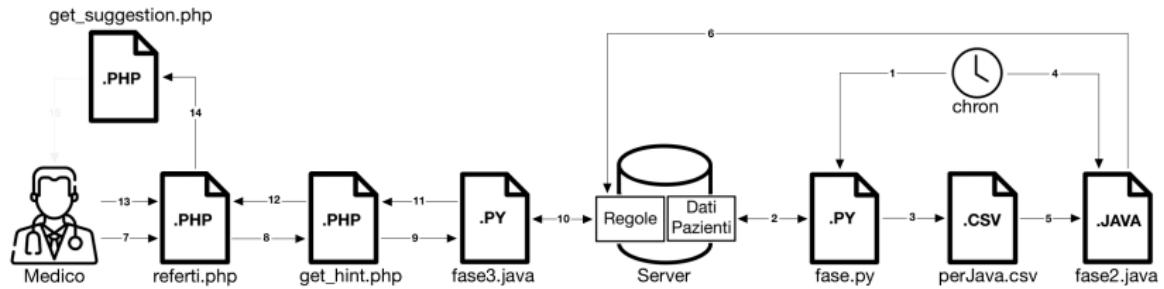
Ciclo?



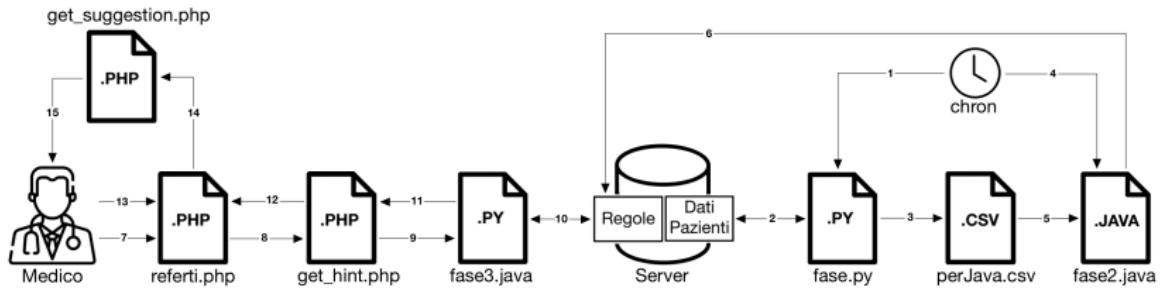
Ciclo?



Ciclo?



Ciclo?



Trasformazione regole

```
(SITUAZ_COLONNA = Osteoporosi) and (OSTEOPOROSI_GRAVE = 1) => CALCIO_SUPP = 1 (132.0/31.0)
(SITUAZ_COLONNA = Osteoporosi) and (TERAPIA_ALTR0 = 1) => CALCIO_SUPP = 1 (245.0/116.0)
(SITUAZ_FEMORE_SN = Osteoporosi) and (TERAPIA_ALTR0 = 1) and (FRATTURA_FAM = 0) => CALCIO_SUPP = 1 (73.0/23.0)
(OSTEOPOROSI_GRAVE = 1) => CALCIO_SUPP = 1 (63.0/18.0)
(SITUAZ_COLONNA = Osteoporosi) and (INTOLLERANZE_0 = 1) => CALCIO_SUPP = 1 (10.0/0.0)
=> CALCIO_SUPP = 0 (1296.0/156.0)
```

Le regole risultanti hanno un implicito ordine di esecuzione. Le regole prese individualmente potrebbero essere non valide.
Risulta necessario ristrutturare le regole.

Trasformazione regole

Le regole ereditano le condizioni negate delle regole precedenti.

if (c_1) return y_1 else if (c_2) return y_2 ⋮ else return y_d	→	if (c_1) return y_1 else if ($c_2 \wedge \neg c_1$) return y_2 else if ($c_3 \wedge \neg c_1 \wedge \neg c_2$) return y_3 ⋮ else if ($\neg c_1 \wedge \dots \wedge \neg c_{d-1}$) return y_d
--	---	--

Performance regole

La performance di una regola si giudica in base al valore di *support* e *confidence*.

	Rilevante	Non rilevante
Affidabile	Decisamente desiderabile	Desiderabile
Non affidabile	Desiderabile	Non desiderabile

Riadattamento regole

Le regole trasformate possono apparire eccessivamente complesse per l'utente finale.

$$r_i : (c_i \wedge \neg c_1 \wedge \dots \wedge \neg c_{i-1}) \rightarrow y_i$$

$$\begin{aligned} r_i : & ((A_{i,1} \text{ op } v_{i,1} \wedge \dots \wedge A_{i,n_i} \text{ op } v_{1,n_i}) \wedge (A_{1,1} \neg \text{op } v_{1,1} \vee \dots \vee \\ & A_{1,n_1} \neg \text{op } v_{1,n_1}) \wedge \dots \wedge (A_{i-1,1} \neg \text{op } v_{i-1,1} \vee \dots \vee \\ & A_{i-1,n_{i-1}} \neg \text{op } v_{i-1,n_{i-1}})) \rightarrow y_i \end{aligned}$$

Le regole trasformate possono apparire eccessivamente complesse per l'utente finale.

Per cui vengono riadattate.

$$r_i : ((A_{i,1} \ op \ v_{i,1} \wedge \dots \wedge A_{i,n_i} \ op \ v_{1,n_i}) \wedge (A_{1,T} \ \neg op \ v_{1,T}) \wedge \dots \wedge (A_{i-1,T} \ \neg op \ v_{i-1,T})) \rightarrow y_i$$

Si rimuovono le proposizioni ridondanti.

$$r_3 : (BMI < 30 \wedge OSTEOPOROSI_GRAVE = 0 \wedge BMI < 25 \wedge OSTEOPOROSI_GRAVE = 0) \rightarrow si$$
$$r_3 : (BMI < 25 \wedge OSTEOPOROSI_GRAVE = 0) \rightarrow si$$

Output utente

Si riconverte dalla procedura di one-hot-encoding.

The diagram illustrates the transformation of a categorical feature 'ALLERGIE' into a binary matrix. On the left, a table has a single column labeled 'ALLERGIE' containing values 'A', 'B,C', and 'C'. An arrow points to the right, where the same data is represented as a matrix with three columns: 'ALLERGIE_0', 'ALLERGIE_1', and 'ALLERGIE_2'. The matrix rows correspond to the categories A, B,C, and C. The 'ALLERGIE_0' column contains values 1, 0, and 0 respectively. The 'ALLERGIE_1' column contains values 0, 1, and 0. The 'ALLERGIE_2' column contains values 0, 0, and 1. Ellipses indicate other rows in the original table and the resulting matrix.

...	ALLERGIE	...
...	A	...
...	B,C	...
...	C	...
...		...

→

...	ALLERGIE_0	ALLERGIE_1	ALLERGIE_2	...
...	1	0	0	...
...	0	0	0	...
...	0	1	1	...
...	0	0	1	...
...				...

$r_5 : (\dots \wedge ALLERGIE_2 = 1 \wedge \dots) \rightarrow si$

$r_5 : (\dots \wedge Il\ campo\ allergie\ contiene\ C \wedge \dots) \rightarrow si$

Output utente

Si suggerisce la terapia

Il campo INTOLLERANZE contiene 'lattosio',
FRAX COLLO FEMORE ≥ 0.9 ,
BMI ≤ 25.3906 ,

OSTEOPOROSI GRAVE: no,
DEFRA < 12,

TOTAL T SCORE > -2.5,
TOTAL Z SCORE > -0.5,
ETA MENOPAUSA > 49

(regola affidabile ma non rilevante)

di tipo calcio citrato

INTOLLERANZE: si
(regola affidabile e rilevante)

Dopo questa fase l'output utente risulta molto più ordinato e leggibile.

Conclusioni

TERAPIA ORMONALE	88%
TERAPIA OSTEOPROTETTIVA	90%
TERAPIA A BASE DI VIT. D	87%
SUPPLEMENTAZIONE CALCIO	89%
SUPPLEMENTAZIONE VIT. D	91%
PRINCIPIO ATTIVO TERAPIA ORMONALE	60%
PRINCIPIO ATTIVO TERAPIA OSTEOPROTETTIVA	45%
PRINCIPIO ATTIVO TERAPIA A BASE DI VIT. D	50%
PRINCIPIO ATTIVO SUPPLEMENTAZIONE CALCIO	65%
PRINCIPIO ATTIVO SUPPLEMENTAZIONE VIT. D	60%

Il sistema progettato appare promettente, per di piu , con il tempo ci si aspetta un continuo miglioramento delle raccomandazioni.

Attualmente si sta lavorando ad una versione piu' matura con numerosi miglioramenti.