

Fuzzy Inference System Professional

Presentazione del software **FisPro**

Corso di studio: Trattamento dell'incertezza nell'informazione

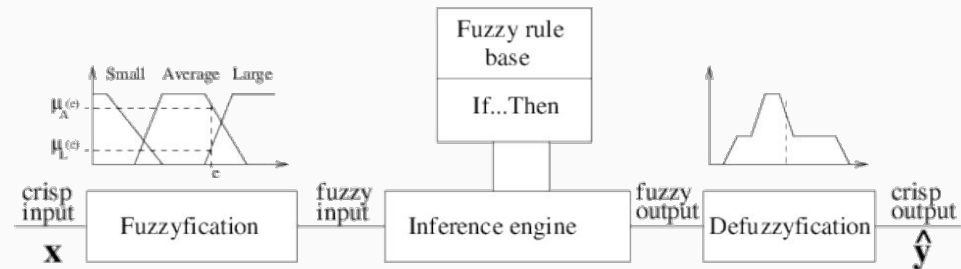
Candidato: Emilio Acciaro



FisPro



- Software per la progettazione di *Fuzzy Inference System* utilizzabili per il *reasoning*
 - Capacità di simulare sistemi fisici e biologici
- Creazione di sistemi a partire da conoscenza specifica
- Sviluppo di sistemi fuzzy a partire dai dati



Elementi chiave di FisPro

- Interpretabilità del sistema garantita ad ogni step:
 - partizionamento delle variabili
 - induzione delle regole
 - semplificazione delle regole
 - ottimizzazione
- Architettura modulare e portabile che garantisce l'indipendenza dalla piattaforma di utilizzo
- Software free e open source, permette:
 - facilità di utilizzo
 - studio
 - cambiamenti
 - miglioramenti

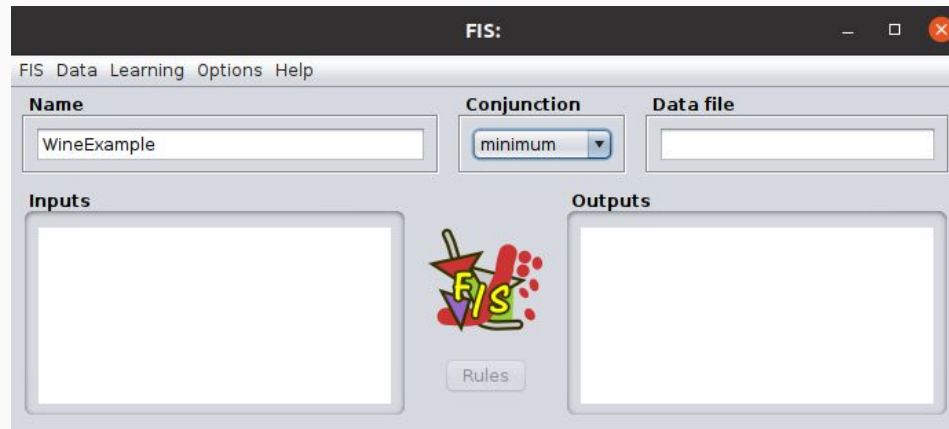
Costruire un FIS basato
su conoscenza:
esempio introduttivo

FIS a due variabili

- Si supponga di voler costruire un sistema che dia in output il prezzo del vino a partire da due variabili di input, il grado e l'invecchiamento.
- Le regole che si ottengono cambiano il prezzo in funzione del grado e dell'invecchiamento.

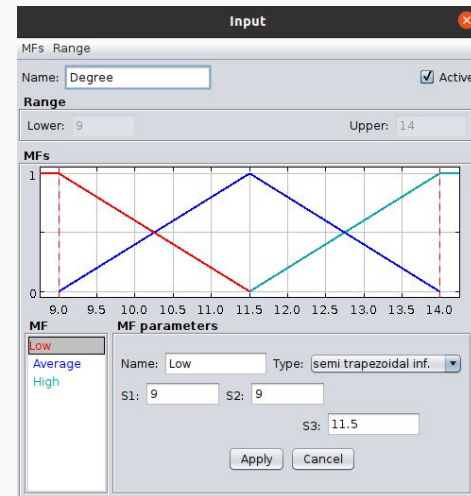
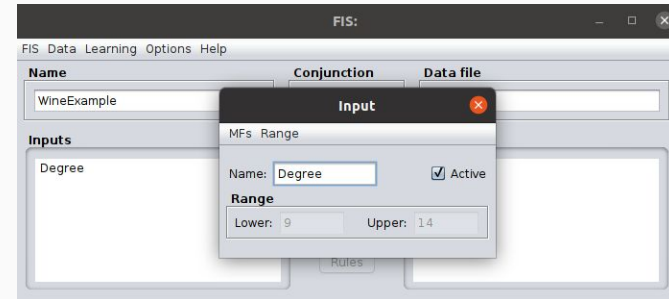
Creazione del sistema

- Per iniziare, creare un FIS selezionando l'opzione *New* dal menu FIS
 - rinominare il nuovo sistema dalla casella editabile visibile nella finestra principale
- Scegliere l'operatore di congiunzione per combinare i valori ottenuti dalle funzioni di appartenenza nella parte antecedente della regola.



Definizione delle variabili di input

- Selezionare l'opzione Inputs -> New input dal menu FIS
 - rinominare con Degree
 - definire il range di input
 - selezionare Range da Input e inserire 9-14
 - Definire una partizione sfruttando l'opzione Regular grid dal menu MF, scegliendo il numero di funzioni di appartenenza corrispondenti al numero di termini linguistici.



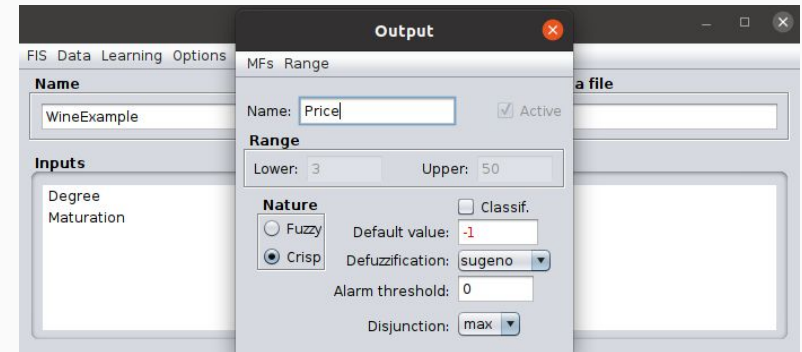
Definizione delle variabili di input

- Ripetere la procedura per creare la variabile maturation, con range di input 2005-2020
- Modificare il nome delle funzioni di appartenenza editando il campo Name, confermare con Apply.
- È possibile modificare il tipo di funzione come anche i vertici, modellando in base alla conoscenza posseduta.



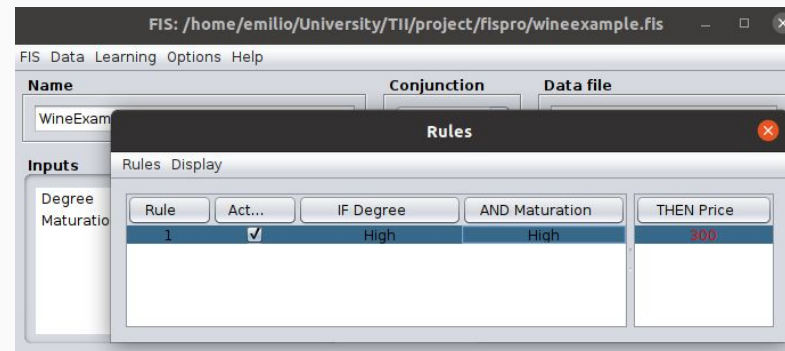
Definizione della variabile di output

- Selezionare l'opzione Outputs -> New output dal menu FIS
 - rinominare con Price
 - definire il range dei valori
 - selezionare Range da Input e inserire 3-50
 - definire la natura dell'output: crisp o fuzzy
 - il valore di default è quello scelto quando nessuna regola è dedotta.
 - i metodi di defuzzificazione disponibili sono: Sugeno e 'max crisp'



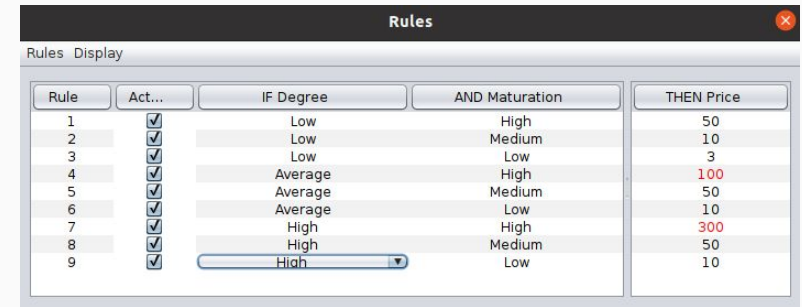
Creazione delle regole

- Premere il pulsante *Rules* dalla finestra principale
 - Dal menu *Rules*, selezionare l'opzione *New Rule*
 - Inserire per ciascuna variabile di input un'etichetta dal menu a tendina a comparsa
 - Inserire un valore per l'output
- Nell'esempio riportato è stata creata una regola tale per cui, se il grado è 'alto' e l'annata è elevata allora il prezzo corrisponde a 300



Creazione delle regole: generazione di tutte le combinazioni

- Le variabili di input per questo esempio sono esclusivamente due.
- Nel caso in cui ci si trovi a lavorare con più variabili, può tornare utile lo strumento *Generate Rules*, selezionabile dal menu *FIS*
- Premendo sul pulsante Rules dal menu principale, compaiono tutte le regole generate.
- Sulla base della propria esperienza/conoscenza è possibile inserire i valori di output corrispondenti



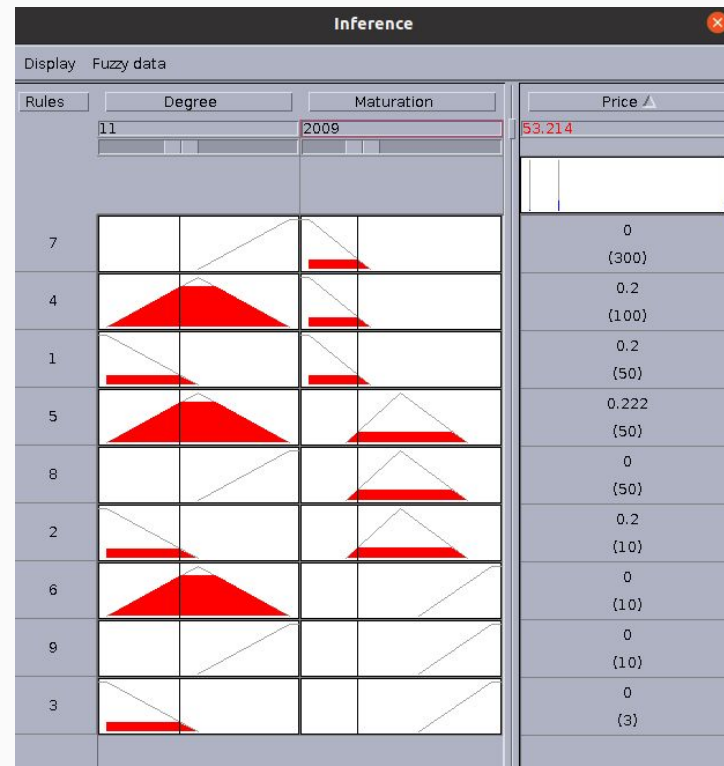
| Rule | Act... | IF Degree | AND Maturation | THEN Price |
|------|-------------------------------------|-----------|----------------|------------|
| 1 | <input checked="" type="checkbox"/> | Low | High | 50 |
| 2 | <input checked="" type="checkbox"/> | Low | Medium | 10 |
| 3 | <input checked="" type="checkbox"/> | Low | Low | 3 |
| 4 | <input checked="" type="checkbox"/> | Average | High | 100 |
| 5 | <input checked="" type="checkbox"/> | Average | Medium | 50 |
| 6 | <input checked="" type="checkbox"/> | Average | Low | 10 |
| 7 | <input checked="" type="checkbox"/> | High | High | 300 |
| 8 | <input checked="" type="checkbox"/> | High | Medium | 50 |
| 9 | <input checked="" type="checkbox"/> | High | Low | 10 |

Meccanismo di inferenza

- L'opzione Infer dal menu FIS mostra graficamente il meccanismo di inferenza
- E' possibile inserire i valori di input direttamente per ciascuna variabile o muovere il cursore per scegliere un valore nel range di appartenenza.
- Per ciascuna regola:
- l'area rossa rappresenta il grado di appartenenza delle variabili di input
- il matching degree si ottiene come combinazione dei livelli di appartenenza presenti nella parte antecedente della regola



- Inserire come valori di input:
 - 11 per la variabile *Degree* e 2009 per la variabile *Maturation*
- Il valore di output per *Price* è crisp; avendo scelto come metodo di defuzzificazione *Sugeno* e come aggregazione la somma, il valore ottenuto dall'inferenza non è altro che una somma pesata delle conclusioni relative ad ogni regola. I pesi in questione corrispondono ai matching degree delle regole.
- Infatti:
 - $0,2 * 100 + 0,2 * 50 + 0,222 * 50 + 0,2 * 10 = 53,214$

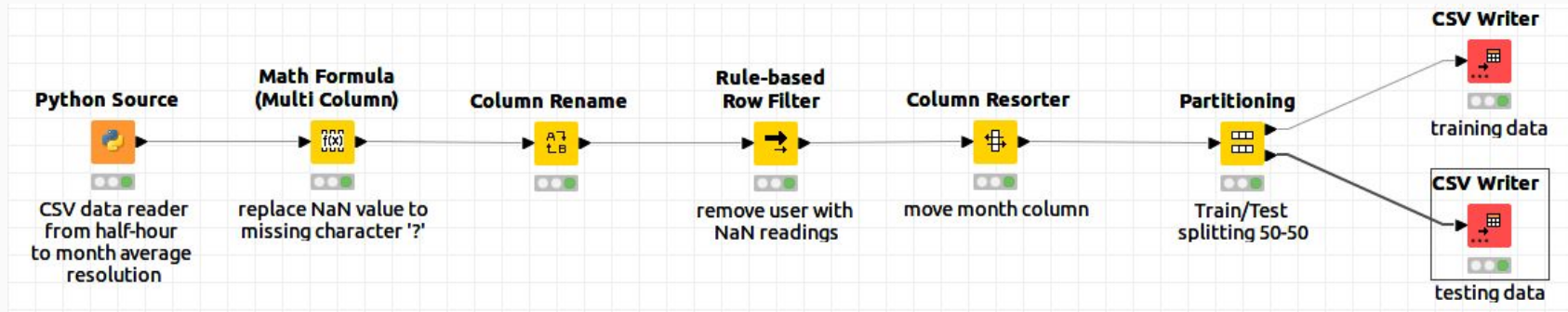


Costruire un FIS a partire dai dati

- Nel dataset a disposizione sono contenute le rilevazioni ogni mezz'ora dei kWh consumati nell'arco di un intero anno per circa tremila utenze
- L'obiettivo è predire i consumi medi per ciascun mese dell'anno considerando esclusivamente i restanti mesi.
- A tale scopo, per ogni mese dell'anno, è stato creato un dataset che abbia come output il mese preso in considerazione.
- Per la fase di sperimentazione, sono stati creati degli script in modo tale da testare i vari metodi messi a disposizione da FisPro

Predizione dei consumi energetici con FisPro

- Le procedure di trattamento del dato sono state condotte con il software *Knime*, il quale permette di sviluppare workflow in maniera grafica, sfruttando i vantaggi della programmazione visuale.



Predizione dei consumi energetici con FisPro

- Nello specifico, sono state condotte le seguenti trasformazioni:
 - Per ridurre il numero di feature, la risoluzione è stata convertita in una media mensile
 - Sono state rimosse le utenze che presentavano valori mancanti
 - Per ciascun mese:
 - riordinamento delle colonne del dataset, per ottenere il mese corrente come ultima variabile
 - split del dataset in due parti (50-50): training e testing

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 0 | 0.526612 | 0.399472 | 0.278366 | 0.204328 | 0.186498 | 0.130743 | 0.150723 | 0.161161 | 0.157917 | 0.173101 | 0.314014 | 0.340036 |
| 1 | 0.479726 | 0.373520 | 0.300380 | 0.231160 | 0.202748 | 0.198989 | 0.190112 | 0.155836 | 0.210272 | 0.192497 | 0.304420 | 0.364181 |
| 2 | 0.305519 | 0.258193 | 0.234994 | 0.209737 | 0.192261 | 0.181058 | 0.176942 | 0.175684 | 0.188124 | 0.192876 | 0.263134 | 0.308794 |
| 3 | 1.285661 | 1.116443 | 0.854362 | 0.663128 | 0.490339 | 0.226480 | 0.219765 | 0.235983 | 0.267137 | 0.465462 | 1.049662 | 1.163352 |
| 4 | 0.004834 | 0.002786 | 0.001255 | 0.058648 | 0.091137 | 0.012347 | 0.030465 | 0.006315 | 0.011681 | 0.006617 | 0.000090 | 0.040966 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 130 | 0.163242 | 0.145512 | 0.082266 | 0.102892 | 0.107786 | 0.100325 | 0.091717 | 0.107362 | 0.109768 | 0.096171 | 0.101394 | 0.079307 |
| 131 | 0.324769 | 0.275754 | 0.246401 | 0.221969 | 0.179239 | 0.168958 | 0.161978 | 0.175829 | 0.136183 | 0.210078 | 0.269145 | 0.309809 |
| 132 | 0.048890 | 0.047057 | 0.045958 | 0.046260 | 0.046463 | 0.069743 | 0.117803 | 0.119480 | 0.149959 | 0.187632 | 0.244601 | 0.208284 |
| 133 | 0.250790 | 0.200663 | 0.119414 | 0.098374 | 0.096826 | 0.085354 | 0.149026 | 0.067454 | 0.081397 | 0.082560 | 0.129751 | 0.119515 |
| 134 | 0.570122 | 0.452715 | 0.399313 | 0.349517 | 0.205803 | 0.123337 | 0.123314 | 0.122279 | 0.166960 | 0.240796 | 0.487420 | 0.572418 |

Training data

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|-------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| count | 135.000000 | 135.000000 | 135.000000 | 135.000000 | 135.000000 | 135.000000 | 135.000000 | 135.000000 | 135.000000 | 135.000000 | 135.000000 | 135.000000 |
| mean | 0.271776 | 0.247581 | 0.212246 | 0.189470 | 0.172943 | 0.156910 | 0.156896 | 0.156067 | 0.165471 | 0.184047 | 0.225990 | 0.256117 |
| std | 0.225064 | 0.197642 | 0.166118 | 0.137263 | 0.124799 | 0.111855 | 0.112502 | 0.114239 | 0.115447 | 0.127042 | 0.183580 | 0.214917 |
| min | 0.004834 | 0.002786 | 0.001255 | 0.035841 | 0.037280 | 0.012347 | 0.026408 | 0.006315 | 0.011681 | 0.006617 | 0.000090 | 0.007388 |
| 25% | 0.127085 | 0.121444 | 0.112830 | 0.100861 | 0.095967 | 0.088013 | 0.091241 | 0.089347 | 0.094538 | 0.107632 | 0.119341 | 0.125598 |
| 50% | 0.199119 | 0.192417 | 0.166047 | 0.155094 | 0.142569 | 0.127183 | 0.128550 | 0.128272 | 0.142358 | 0.149398 | 0.165616 | 0.189625 |
| 75% | 0.308528 | 0.284866 | 0.259994 | 0.226956 | 0.204337 | 0.190807 | 0.189150 | 0.185500 | 0.196371 | 0.215123 | 0.279677 | 0.316781 |
| max | 1.285661 | 1.116443 | 1.028802 | 0.830233 | 0.863786 | 0.770507 | 0.865174 | 0.835168 | 0.722331 | 0.762964 | 1.049662 | 1.163352 |

Statistiche per ciascuna feature

Esempio con FisPro GUI

- Selezionando l'opzione *Open* dal menu *Data* è possibile aprire il file contenente i dati
- Il primo step consiste nel partizionare le variabili
- In FisPro esistono essenzialmente due approcci:
 - il primo consiste nel creare un FIS senza regole, da cui è possibile scegliere il numero di MF per ciascuna variabile, il metodo di partizionamento, la tipologia di output (crisp o fuzzy) e infine il metodo di defuzzificazione
 - il secondo consiste nell'utilizzare il metodo noto come *Hierarchical Fuzzy Partitioning*, che include sia una fase di partizionamento ma anche una fase di selezione delle regole

Generazione di un FIS senza regole

- Selezionare *Partitions* -> *Generate FIS without rules* dal menu principale Learning
- È possibile scegliere:
 - il numero di MF per ciascuna variabile
 - la gerarchia di partizionamento
 - hfp
 - k-means
 - regular
 - la tipologia di output, crisp o fuzzy
 - il metodo di defuzzificazione
 - area
 - mean max
 - sugeno

Generate FIS without rules

Inputs

Number of MFs

3 All inputs

| | | | |
|-----|---|-----|---|
| Jan | 3 | Feb | 3 |
| Mar | 3 | Apr | 3 |
| May | 3 | Jun | 3 |
| Jul | 3 | Aug | 3 |
| Sep | 3 | Oct | 3 |
| Nov | 3 | | |

Hierarchy: hfp

initial partition

☐ Nb of groups: 100

☒ Tolerance threshold: 0.01

Output

Conjunction: minimum

☐ Fuzzy

☒ Crisp

Number of MFs: 3

Hierarchy: regular

initial partition

☒ Nb of groups: 100

☐ Tolerance threshold: 0.01

Defuzzification: sugeno

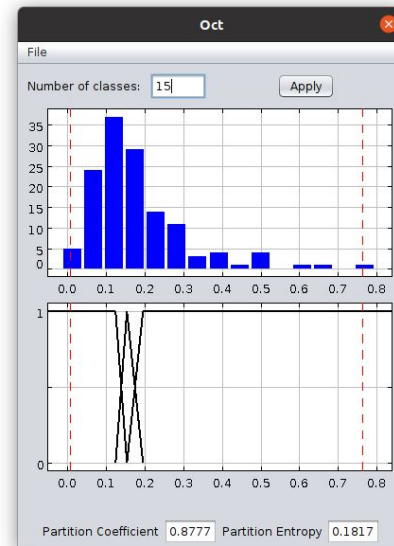
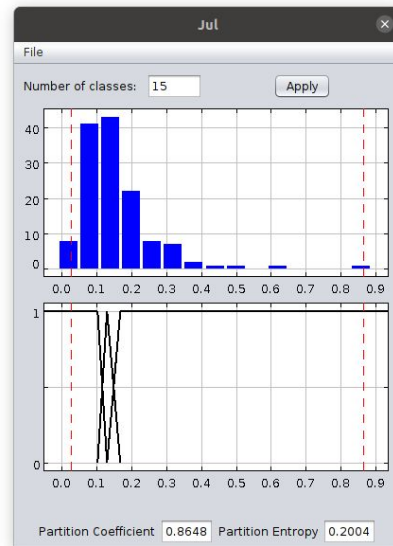
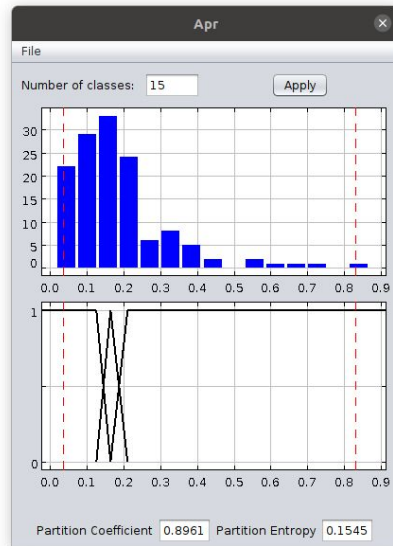
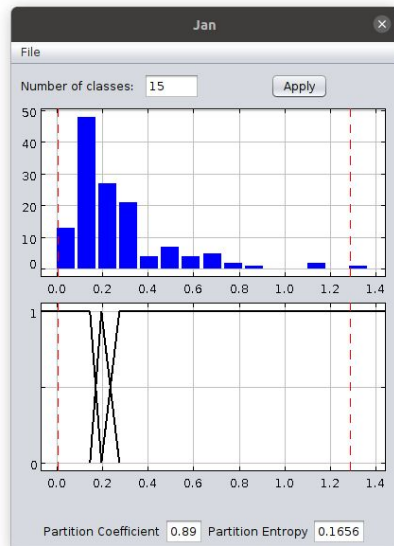
Disjunction: sum

☐ Classif.

Apply Cancel

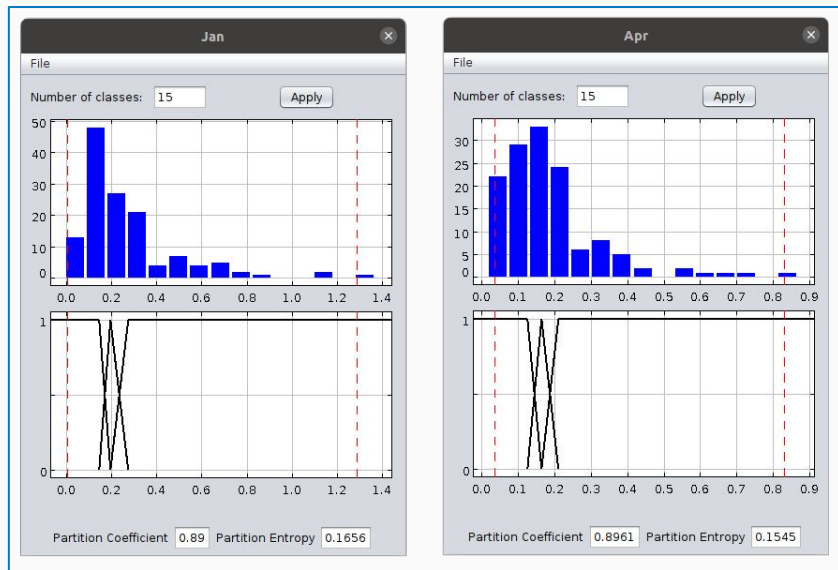
Generazione di un FIS senza regole

- Selezionando l'opzione *View* dal menu *Data* è possibile visualizzare la distribuzione dei dati per ogni variabile, insieme al partizionamento ottenuto per ciascuna di esse.

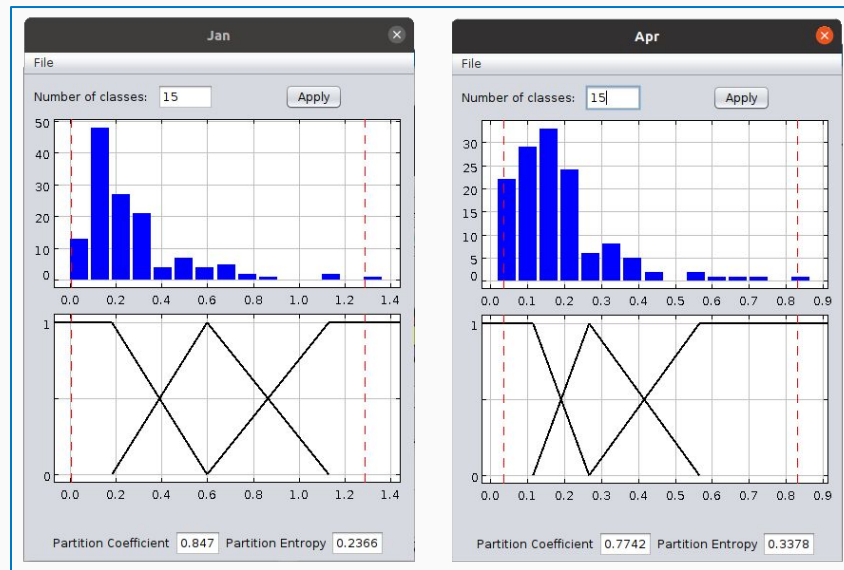


Generazione di un FIS senza regole

- A seguire un confronto fra il partizionamento ottenuto con il K-means e quello ottenuto con il metodo HFP.
- È possibile notare come il partizionamento ottenuto con quest'ultimo sia migliore rispetto a quello ottenuto con il K-means. Dal punto di vista qualitativo, è possibile basarsi sui valori di *Partition Coefficient* e *Partition Entropy*



Partizionamento con HFP



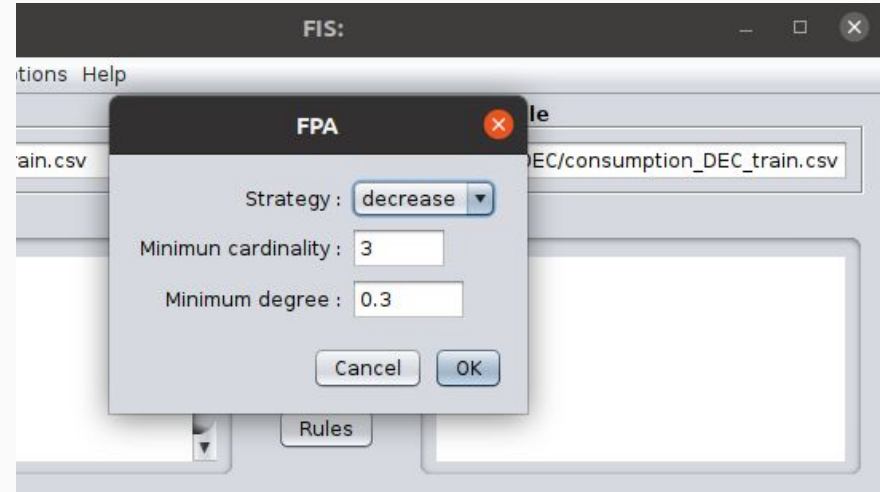
Partizionamento con K-means

Induzione delle regole con il metodo FPA

- Utilizzando il metodo FPA è possibile generare le regole e impostare le relative conclusioni.
- Il metodo FPA, che sta per *Fast Prototype algorithm*, è una tecnica efficiente che permette di inizializzare o aggiornare le conclusioni delle regole a partire dai dati.
- Le conclusioni sono calcolate utilizzando i valori osservati rispetto ad un sottoinsieme degli esempi. Per ciascuna regola si utilizza un diverso sottoinsieme di esempi.
- L'output per questo esempio applicativo è un valore continuo, pertanto le conclusioni sono inizializzate con la somma dei valori di output osservati nel sottoinsieme considerato.
- Per ciascun output osservato abbiamo un peso che corrisponde al matching degree

Induzione delle regole con il metodo FPA

- Selezionare *Rule induction* -> *FPA* dal menu Learning
- I parametri riguardano gli aspetti relativi alla selezione del sottoinsieme di esempi per ciascuna regola.



Induzione delle regole con il metodo FPA

| Rules | | | | | | | | | | | | | |
|---------------|-------------------------------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Rules Display | | | | | | | | | | | | | |
| Rule | Act... | IF J... | AN... | AN... | AN... | AN... | AN... | AN... | AN... | AN... | AN... | AN... | TH... |
| 1 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | 0.086 |
| 2 | <input checked="" type="checkbox"/> | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | 0.517 |
| 3 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF1 | MF1 | 0.177 |
| 4 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF1 | 0.135 |
| 5 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | 0.191 |
| 6 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF2 | MF2 | 0.204 |
| 7 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | MF1 | MF1 | 0.136 |
| 8 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | 0.115 |
| 9 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF3 | MF3 | MF2 | MF2 | 0.208 |
| 10 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF3 | MF3 | MF3 | MF2 | MF2 | 0.206 |
| 11 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | MF1 | MF1 | MF1 | 0.141 |
| 12 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF1 | 0.128 |
| 13 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF2 | MF2 | 0.176 |
| 14 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF2 | MF2 | MF2 | 0.192 |
| 15 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF2 | MF3 | MF2 | MF2 | MF2 | 0.205 |
| 16 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF2 | 0.209 |
| 17 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF1 | MF2 | MF1 | MF1 | MF1 | 0.17 |
| 18 | <input checked="" type="checkbox"/> | MF2 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | 0.262 |
| 19 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | 0.162 |
| 20 | <input checked="" type="checkbox"/> | MF3 | MF2 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | 0.338 |
| 21 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF3 | MF3 | MF2 | 0.212 |
| 22 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF2 | MF2 | MF2 | 0.186 |
| 23 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | MF2 | MF2 | MF1 | MF1 | 0.182 |
| 24 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF3 | MF2 | MF2 | 0.179 |
| 25 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF1 | MF1 | MF1 | MF1 | 0.115 |
| 26 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | MF2 | MF2 | MF1 | MF1 | 0.169 |
| 27 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | 0.211 |
| 28 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF3 | MF2 | MF2 | 0.176 |
| 29 | <input checked="" type="checkbox"/> | MF2 | MF1 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF2 | MF2 | 0.176 |
| 30 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF3 | MF2 | 0.204 |
| 31 | <input checked="" type="checkbox"/> | MF1 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF2 | MF2 | MF2 | 0.177 |
| 32 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF1 | MF2 | 0.146 |
| 33 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF2 | MF1 | 0.144 |
| 34 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF3 | MF2 | MF2 | MF3 | 0.211 |
| 35 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF3 | MF2 | MF2 | 0.177 |
| 36 | <input checked="" type="checkbox"/> | MF1 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF3 | MF2 | MF2 | 0.178 |
| 37 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF2 | MF2 | MF1 | MF1 | MF1 | 0.171 |
| 38 | <input checked="" type="checkbox"/> | MF2 | MF1 | MF2 | MF2 | MF2 | MF3 | MF2 | MF2 | MF3 | MF2 | MF2 | 0.176 |
| 39 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | MF2 | MF1 | 0.144 |

Performance

Results

Out... Blank thres

☒ Links between rules and items

☐ Dec

Results (Blank thres=0.1)

Error

PI

RMSE

MAE

Coverage %

Max error

Induzione delle regole con il metodo FPA

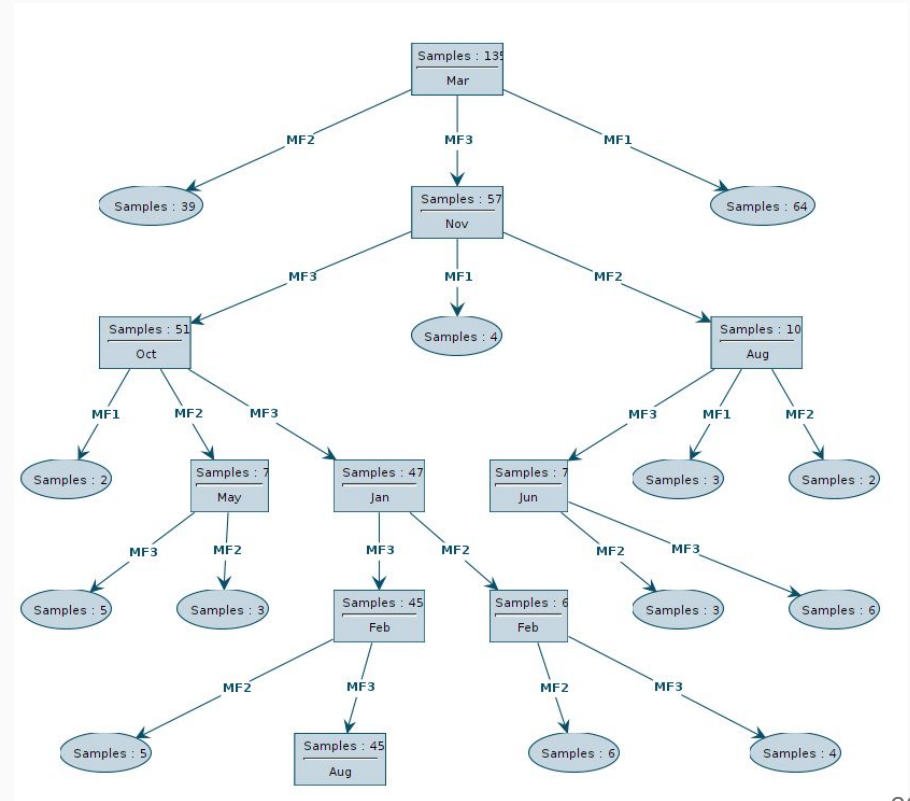
- Ai fini della sperimentazione, la medesima procedura è stata condotta da linea di comando considerando, questa volta, tutti i dataset (che variano in base alla sigla del mese).
- I programmi utilizzati sono i seguenti:
 - **hfpsr** "consumption_DEC_train.csv" "3 3 3 3 3 3 3 3 3 3 3" 1 0.01 3 1 MeanMax sum 0.01
 - **genrules** "consumption_DEC_train.csv-sr.fis"
 - **fpa** "config.fis" "consumption_DEC_train.csv" -f"output.fis" -t"consumption_DEC_test.csv"
 - **perf** "output.fis" "consumption_DEC_test.csv"
- Il primo programma genera le partizioni per le variabili di input e la variabile di output sfruttando come metodo HFP, con metodo di defuzzificazione MeanMax
- Il secondo programma genera tutte le possibili regole, impostando come conclusione il valore di default 1
- Il terzo programma richiama il metodo FPA che ha come obiettivo la generazione delle conclusioni per ogni regola
- Il quarto programma valuta i risultati rispetto al testing set fornendo in output gli indicatori di qualità

Induzione delle regole con i Fuzzy decision Tree

- Un metodo alternativo per l'induzione delle regole sono i Fuzzy decision trees
- Rappresentano un'estensione dei classici alberi di decisione
- Sono costituiti da un nodo radice e da una serie in cascata di ulteriori nodi
- I nodi terminali sono chiamati nodi foglia. Ogni nodo corrisponde ad uno split sui valori di una singola variabile
- Il path dal nodo radice ad un nodo foglia può essere facilmente interpretato come regola di decisione.
- Sono disponibili operazioni di potatura, garantendo la possibilità di ottenere performance migliori così come una più facile interpretabilità dell'albero
- L'algoritmo proposto in FisPro è basato su una implementazione fuzzy dell'algoritmo ID3

Induzione delle regole con i Fuzzy decision Tree

- Dalla GUI di FisPro è possibile costruire fuzzy decision tree in maniera semplice e intuitiva
- Inoltre il programma mette a disposizione tool grafici per la visualizzazione dell'albero di decisione risultante.



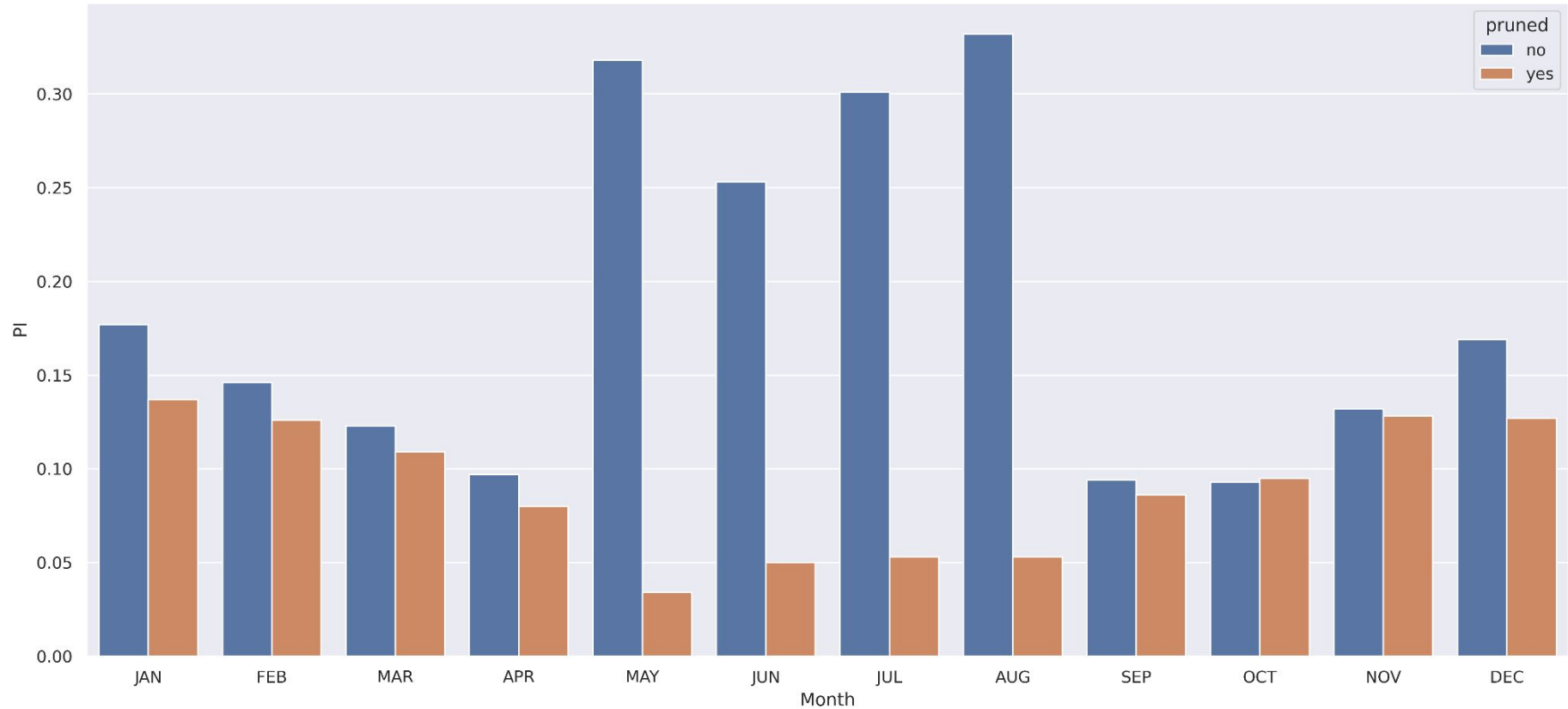
Induzione delle regole con i Fuzzy decision Tree

- Dovendo eseguire una sperimentazione, la procedura è stata condotta da linea di comando considerando, questa volta, tutti i dataset (che variano in base alla sigla del mese).
- I programmi utilizzati sono i seguenti:
 - **hfpsr** "consumption_DEC_train.csv" "3 3 3 3 3 3 3 3 3 3 3 3" 1 0.01 3 1 MeanMax sum 0.01
 - **fistree** "consumption_DEC_train.csv-sr.fis" "consumption_DEC_train.csv" -p1 -v"consumption_DEC_test.csv" -z"consumption_DEC_test.csv"
- Il primo programma genera le partizioni per le variabili di input e la variabile di output sfruttando come metodo HFP, con metodo di defuzzificazione MeanMax
- Il secondo programma richiama l'algoritmo proposto di Fuzzy decision tree per l'induzione delle regole.

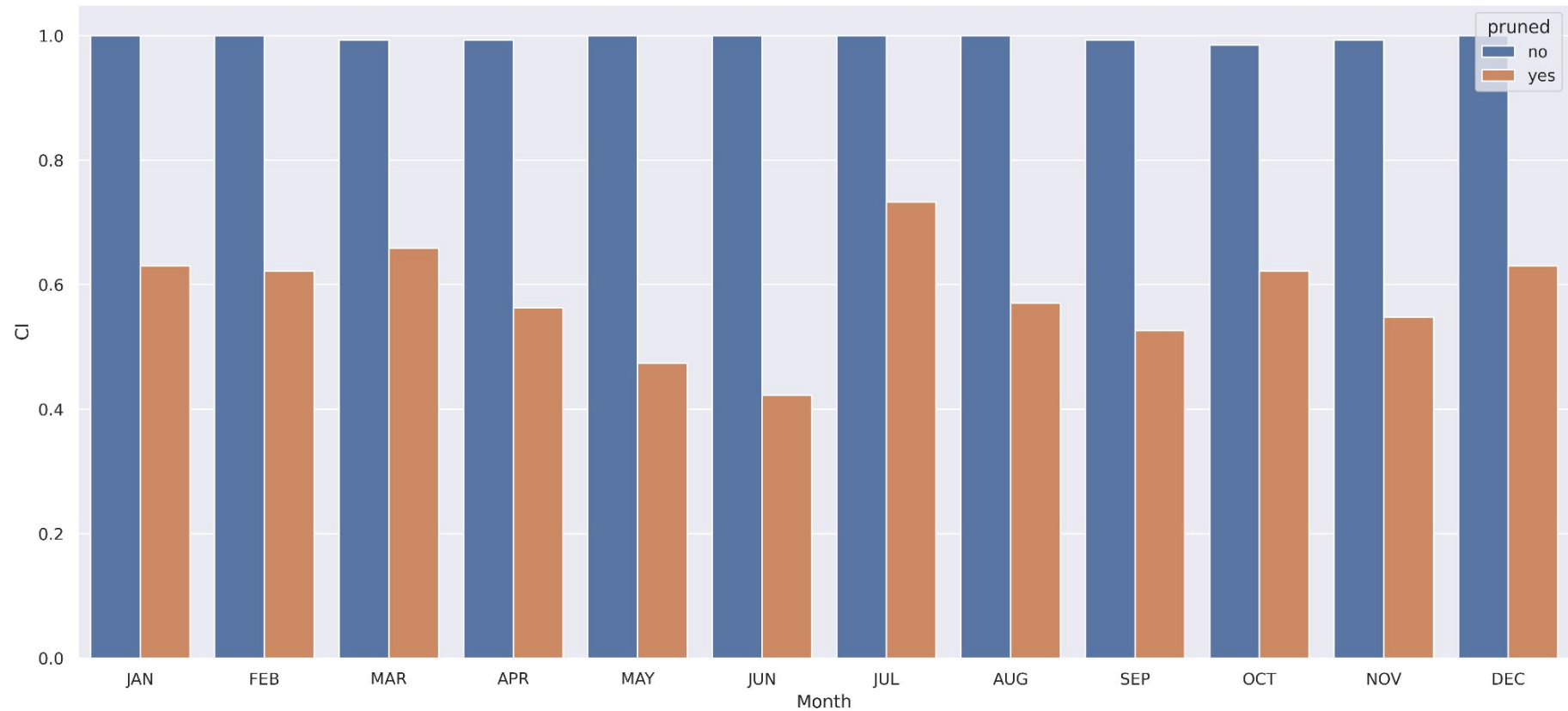
Induzione delle regole con i Fuzzy decision Tree

- Sono state condotte delle analisi intermedie rispetto ai risultati ottenuti con i Fuzzy decision tree
- In particolare, per ciascun dataset, sono mantenuti due alberi di decisione: l'albero originario e quello potato
- A seguire vengono riportati i grafici, considerando:
 - performance
 - coverage
 - numero di regole
- Sulla base dei risultati emersi, sono stati scelti gli alberi potati da mettere a confronto con le altre metodologie di induzione

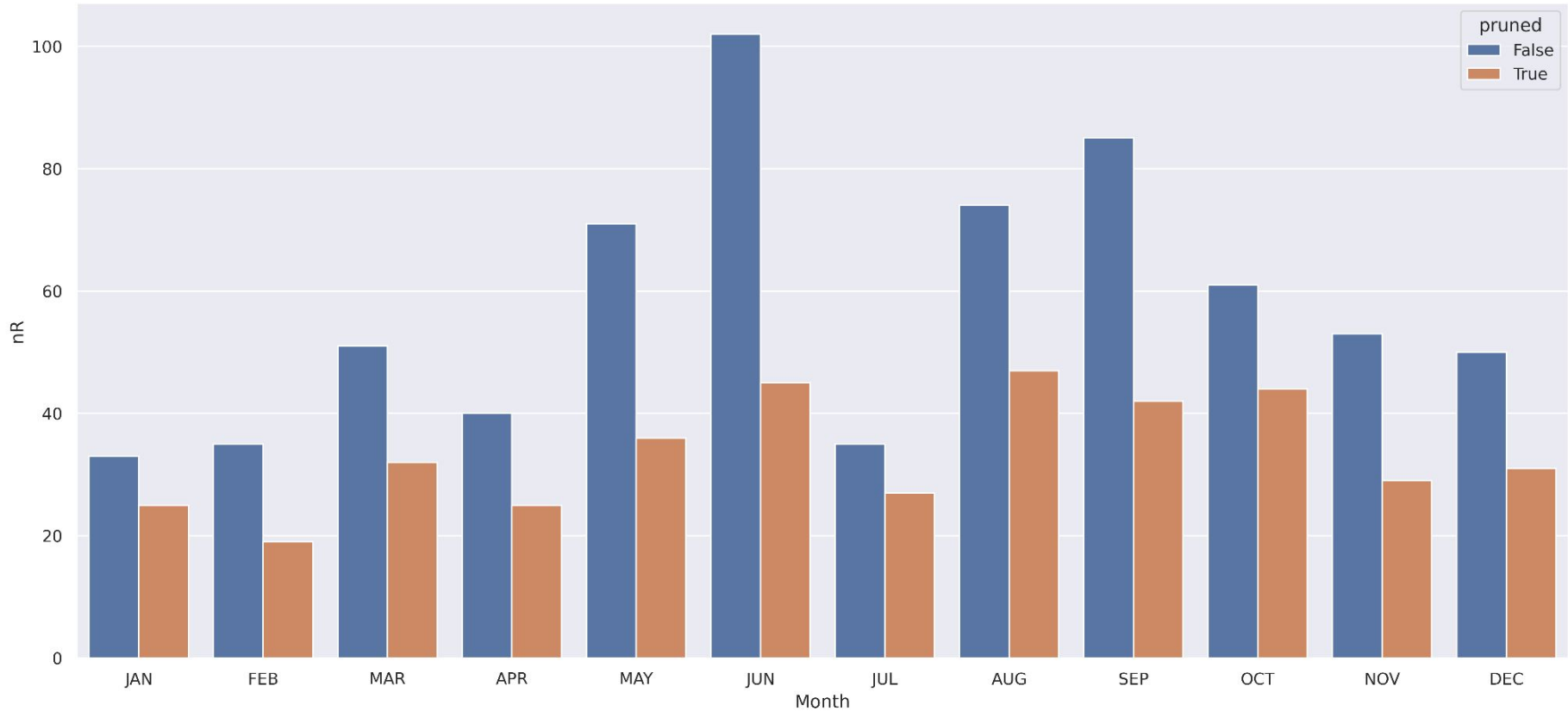
Performance index



Coverage index



Complessità del FIS rispetto al numero di regole



Induzione delle regole con l'algoritmo 'Wang & Mendel'

- Un ulteriore metodo, che permette l'induzione delle regole, è l'algoritmo Wang & Mendel
- Rispetto al metodo originario, per utilizzare questo algoritmo in FisPro è necessario predisporre le partizioni fuzzy.
- Per farlo, è sufficiente generare le partizioni attraverso l'opzione Generate FIS without rules, come precedentemente visto.
- L'algoritmo inizia generando una regola per ogni coppia nel training set.

IF x_1 is A_1^i AND x_2 is A_2^i ... AND x_p is A_p^i THEN y is C^i

- Gli insiemi fuzzy A_j^i sono quelli per cui il grado di appartenenza rispetto a x_j^i è massimo al variare della variabile j per la coppia i .
- È opportuno indicare come tale procedura non sia particolarmente adatta per output crisp rappresentati valori continui.

Induzione delle regole con l'algoritmo 'Wang & Mendel'

- È possibile utilizzare l'algoritmo *Wang & Mendel* da GUI, ma dovendo eseguire una sperimentazione, la procedura è stata condotta da linea di comando considerando, questa volta, tutti i dataset (che variano in base alla sigla del mese).
- I programmi utilizzati sono i seguenti:
 - **hfpsr** "consumption_DEC_train.csv" "3 3 3 3 3 3 3 3 3 3 3 3" 1 0.01 3 1 MeanMax sum 0.01
 - **wm** "consumption_DEC_train.csv-sr.fis" "consumption_DEC_train.csv" -o"output.fis"
 - **perf** "output.fis" "consumption_DEC_test.csv"
- Il primo programma genera le partizioni per le variabili di input e la variabile di output sfruttando come metodo HFP, con metodo di defuzzificazione MeanMax
- Il secondo richiama l'algoritmo Wang & Mendel per l'induzione delle regole.
- Il terzo programma valuta i risultati rispetto al testing set fornendo in output gli indicatori di qualità

- L'algoritmo OLS, che sta per Orthogonal Least Squares, è un ulteriore metodo di induzione messo a disposizione dal tool FisPro.
- Questo metodo trasforma ogni esempio in una regola fuzzy, selezionando in un secondo momento le più importanti con un criterio basato sui minimi quadrati, come la regressione lineare e la ortogonalizzazione di Gram-Schmidt.
- Successivamente vengono ottimizzate le conclusioni per le regole scelte
- In FisPro è possibile utilizzare l'algoritmo OLS passando direttamente i dati di training oppure passando un FIS con le opportune partizioni fuzzy.

Partizionamento e induzione tramite algoritmo OLS

- L'algoritmo originariamente proposto prevedeva la generazione di funzioni di membership Gaussiane per ogni valore di variabile relativa ad un data point.
- Successivamente è condotto un clustering per limitare il numero di funzione di membership.
- È importante notare come le partizioni ottenute non siano standardizzate.
- L'algoritmo, ai fini della sperimentazione, è stato utilizzato da linea di comando.

```
o ols          "consumption_DEC_train.csv"      -q10      -p"consumption_DEC_test.csv"
  -b"consumption_DEC_test.csv"

o perf         "consumption_DEC_train.csv.fis"    "consumption_DEC_test.csv"ols
  "consumption_DEC_train.csv"      -q10      -p"consumption_DEC_test.csv"
  -b"consumption_DEC_test.csv"
```

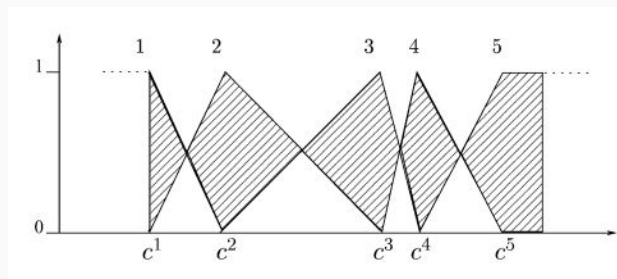
- Il primo programma genera il partizionamento delle variabili e l'induzione delle regole
- Il secondo programma valuta i risultati rispetto al testing set fornendo in output gli indicatori di qualità

Hierarchical Fuzzy Partitioning

- Il metodo HFP rappresenta un caso particolare rispetto agli altri algoritmi.
- L'approccio si ispira a due metodi differenti di clustering: quello gerarchico e quello basato sulla logica fuzzy
- L'idea chiave è ottenere partizioni fuzzy basandosi sull'aggregazione di insiemi fuzzy
- Per l'aggregazione degli insiemi fuzzy, sono sfruttate due misure: la distanza interna che intercorre tra punti che appartengono parzialmente allo stesso fuzzy set; la distanza esterna che intercorre tra punti che appartengono parzialmente a fuzzy set distinti
- In generale, due punti che appartengono prevalentemente allo stesso fuzzy set sono considerati più vicini rispetto ad altri che appartengono a fuzzy set distinti.

Hierarchical Fuzzy Partitioning

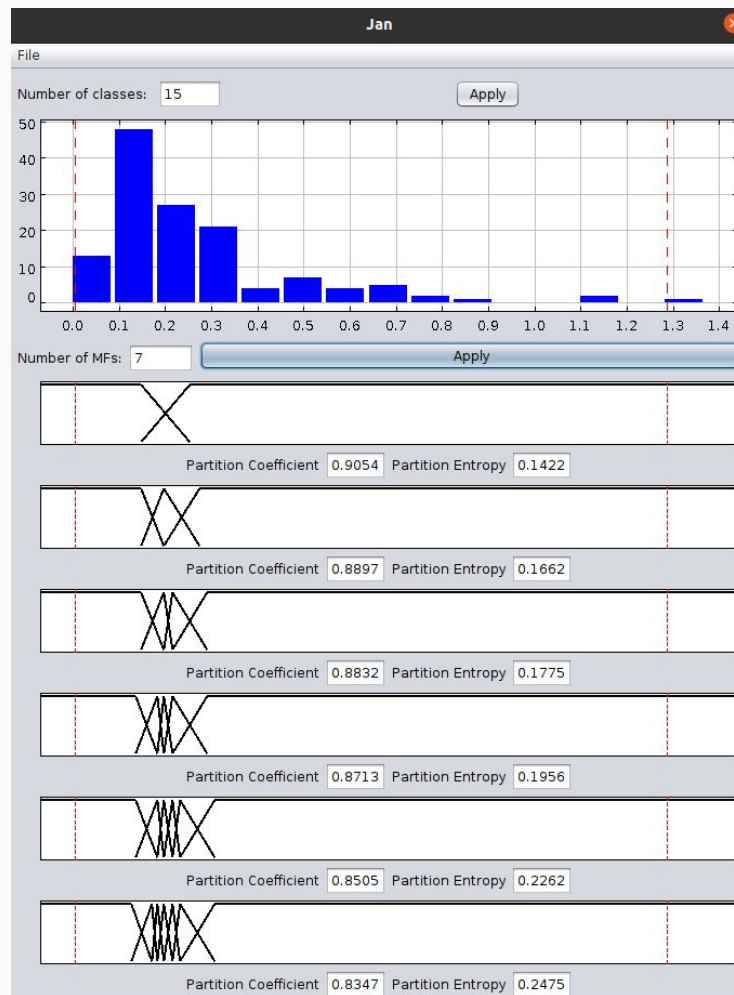
- Per valutare la validità di una partizione fuzzy si utilizza un indice di omogeneità della densità dei fuzzy set.



- L'omogeneità della densità è definita come deviazione standard della densità fra tutti gli insiemi fuzzy della partizione
- Una partizione stabile dovrebbe essere omogenea e quindi avere una bassa deviazione standard

Hierarchical Fuzzy Partitioning

- Una volta generate le partizioni, è possibile visualizzare la sequenza di partizionamento per ciascuna variabile.
- Dal menu *Learning*, selezionare *Partitions* -> *HFP MF* -> *View vertices*.
- È possibile selezionare la variabile di cui si vuole analizzare il partizionamento al variare del numero di MF.



Hierarchical Fuzzy Partitioning

- Una volta ottenute le partizioni fuzzy per ogni dimensione, queste vengono utilizzate in una procedura di raffinamento basata su gerarchie di partizioni fuzzy a dimensione crescente
- Tale procedura è un algoritmo iterativo che ha come obiettivo la selezione delle variabili o degli insiemi fuzzy da introdurre nel FIS
- Un buon FIS consiste in un compromesso tra complessità, data dal numero di regole, e accuratezza
- Il FIS iniziale è quello più semplice, con una sola partizione per variabile
- A seguire si generano sistemi inferenziali temporanei, dove ciascuno di essi corrisponde all'aggiunta di un insieme fuzzy in una specifica dimensione rispetto al FIS iniziale
- Il risultato della procedura è una sequenza di FIS a complessità crescente

- Come anche per gli altri algoritmi, il metodo HFP è utilizzabile facilmente dalla GUI, ma dovendo eseguire una sperimentazione, il metodo è stato lanciato da linea di comando considerando, questa volta, tutti i dataset (che variano in base alla sigla del mese).
- I programmi utilizzati sono i seguenti:
 - **hfpsr** "consumption_DEC_train.csv" "3 3 3 3 3 3 3 3 3 3 3" 1 0.01 3 1 MeanMax sum 0.01
 - **hfpselect** "consumption_DEC_train.csv" "consumption_DEC_train.csv.hfp" -b0.7 -e2 -m0.3 -l"consumption_DEV_train.csv.vertex" -v"consumption_DEC_test.csv"
- Il primo programma genera le partizioni per le variabili di input e la variabile di output sfruttando come metodo HFP, con metodo di defuzzificazione MeanMax
- Il secondo programma sfrutta le partizioni precedentemente generate per creare un file contenente l'insieme dei FIS mantenuti. Per validare i risultati ottenuti è utilizzato il dataset di testing.

Hierarchical Fuzzy Partitioning

- Una volta creato il file di configurazione HFP e generato il file contenente i vertici per ogni partizionamento, è possibile generare un FIS scegliendo il numero di MF per ciascuna variabile.
- Dal menu *Learning*, selezionare *Rule induction -> HFP FIS*

Generate FIS

HFP File: /home/emilio/University/Til/project/fisproGUI-example/DEC/consumption_dec_train.csv.hfp

Vertices file: /home/emilio/University/Til/project/fisproGUI-example/DEC/consumption_DEC_train.csv.vertices

Number of MFs

| | | | |
|-----|---|-----|---|
| Jan | 3 | Feb | 3 |
| Mar | 3 | Apr | 3 |
| May | 3 | Jun | 3 |
| Jul | 3 | Aug | 3 |
| Sep | 3 | Oct | 3 |
| Nov | 3 | | |

Rule induction method

☒ FPA
☐ Wang & Mendel

Cardinality strategy: decrease

Minimum matching degree: 0.3

Minimum cardinality: 3

Minimum cumulated weight: 0

Apply Cancel

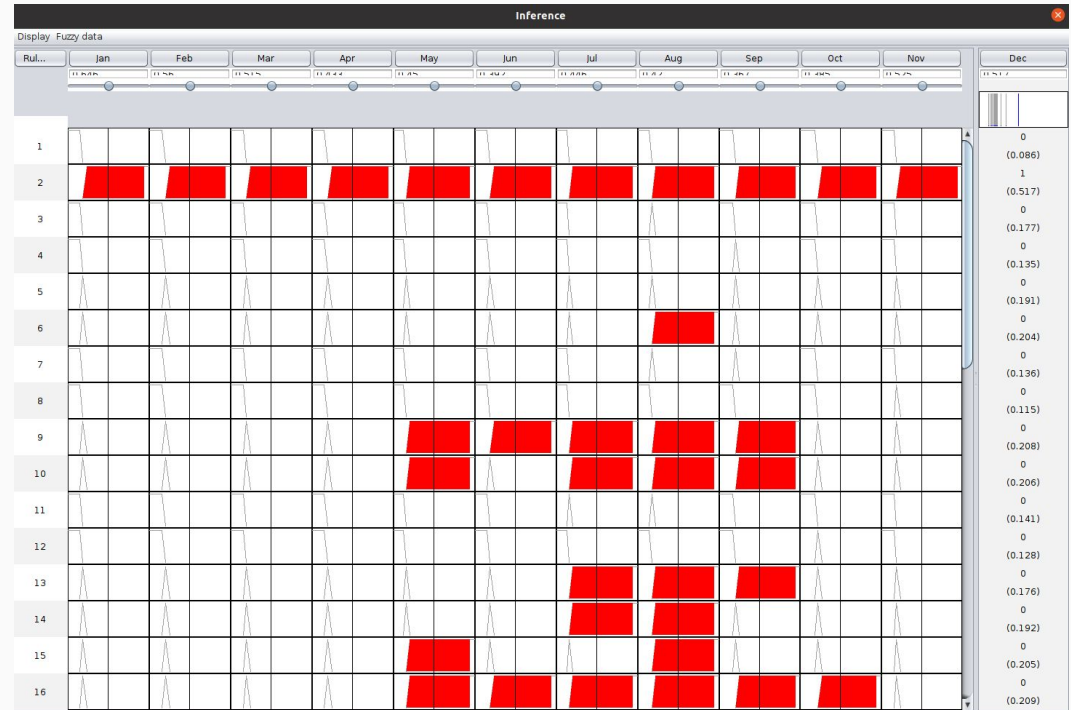
Hierarchical Fuzzy Partitioning

- Di seguito le regole indotte tramite il metodo HFP FIS

| Rules | | | | | | | | | | | | | |
|---------------|-------------------------------------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|
| Rules Display | | | | | | | | | | | | | |
| Rule | Act... | IF Jan | AND Feb | AND Mar | AND Apr | AND May | AND Jun | AND Jul | AND Aug | AND Sep | AND Oct | AND Nov | THEN Dec |
| 1 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | 0.086 |
| 2 | <input checked="" type="checkbox"/> | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | 0.517 |
| 3 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | 0.177 |
| 4 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | 0.135 |
| 5 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | 0.191 |
| 6 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF2 | MF2 | 0.204 |
| 7 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | MF1 | MF1 | 0.136 |
| 8 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | 0.115 |
| 9 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF3 | MF3 | MF2 | MF2 | 0.208 |
| 10 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF3 | MF3 | MF3 | MF2 | MF2 | 0.206 |
| 11 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | MF1 | MF1 | MF1 | 0.141 |
| 12 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | 0.128 |
| 13 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF2 | MF2 | 0.176 |
| 14 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF2 | MF2 | MF2 | 0.192 |
| 15 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF2 | MF3 | MF2 | MF2 | MF2 | 0.205 |
| 16 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF2 | 0.209 |
| 17 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF1 | MF2 | MF1 | MF1 | MF1 | 0.17 |
| 18 | <input checked="" type="checkbox"/> | MF2 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | 0.262 |
| 19 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | 0.162 |
| 20 | <input checked="" type="checkbox"/> | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | MF3 | 0.338 |
| 21 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF3 | MF3 | MF2 | 0.212 |
| 22 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF2 | MF2 | MF2 | 0.186 |
| 23 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | MF2 | MF2 | MF2 | MF1 | MF1 | 0.182 |
| 24 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF3 | MF2 | MF2 | 0.179 |
| 25 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF1 | MF1 | MF1 | MF1 | 0.115 |
| 26 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | MF2 | MF1 | MF1 | 0.169 |
| 27 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF2 | MF3 | MF3 | MF3 | 0.211 |
| 28 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF2 | MF2 | 0.176 |
| 29 | <input checked="" type="checkbox"/> | MF2 | MF1 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF2 | MF2 | 0.176 |
| 30 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF3 | MF2 | 0.204 |
| 31 | <input checked="" type="checkbox"/> | MF1 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF2 | MF2 | 0.177 |
| 32 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF1 | MF2 | 0.146 |
| 33 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF2 | MF1 | 0.144 |
| 34 | <input checked="" type="checkbox"/> | MF2 | MF2 | MF2 | MF2 | MF3 | MF3 | MF3 | MF3 | MF3 | MF2 | MF3 | 0.211 |
| 35 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF3 | MF2 | MF2 | 0.177 |
| 36 | <input checked="" type="checkbox"/> | MF1 | MF2 | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF3 | MF2 | MF2 | 0.178 |
| 37 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF2 | MF1 | MF2 | MF2 | MF1 | MF1 | MF1 | 0.171 |
| 38 | <input checked="" type="checkbox"/> | MF2 | MF1 | MF2 | MF2 | MF2 | MF2 | MF3 | MF2 | MF2 | MF2 | MF2 | 0.176 |
| 39 | <input checked="" type="checkbox"/> | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF1 | MF2 | MF2 | MF2 | MF1 | 0.144 |

Hierarchical Fuzzy Partitioning

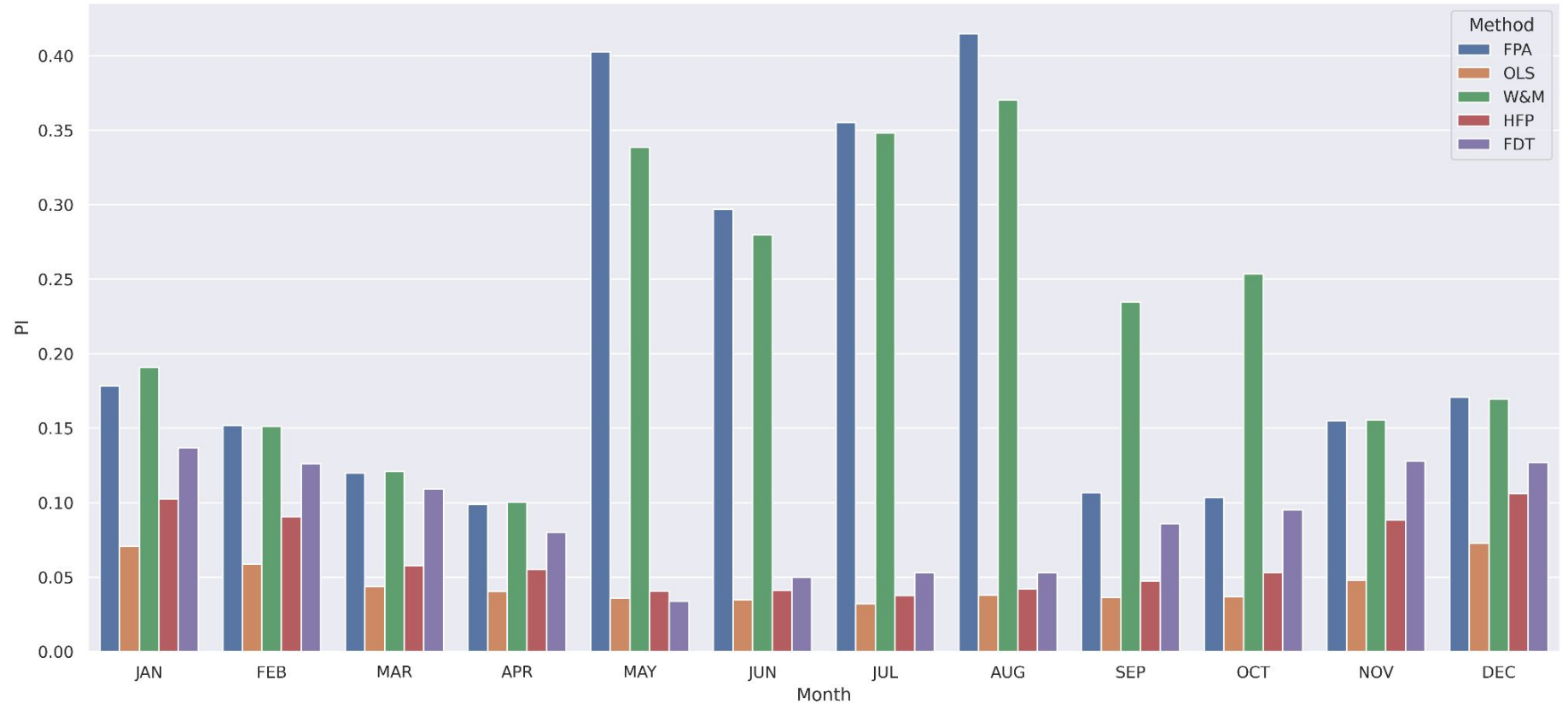
- Dal menu *FIS*, selezionare *Infer* per visualizzare le regole indotte e interagire cambiando i valori per le variabili di input



- Il file di training utilizzato, "consumption_DEC_train.csv" , si contraddistingue rispetto agli altri file di training dalla dicitura **DEC**. Analoghe considerazioni per i file di testing
- Sfruttando uno script scritto in bash, la procedura di sperimentazione è stata resa del tutto automatizzata.
- I risultati migliori per ciascun dataset, validati rispetto ai rispettivi file di testing, sono stati accorpati in un unico file, così da poterli confrontare con le altre metodologie.

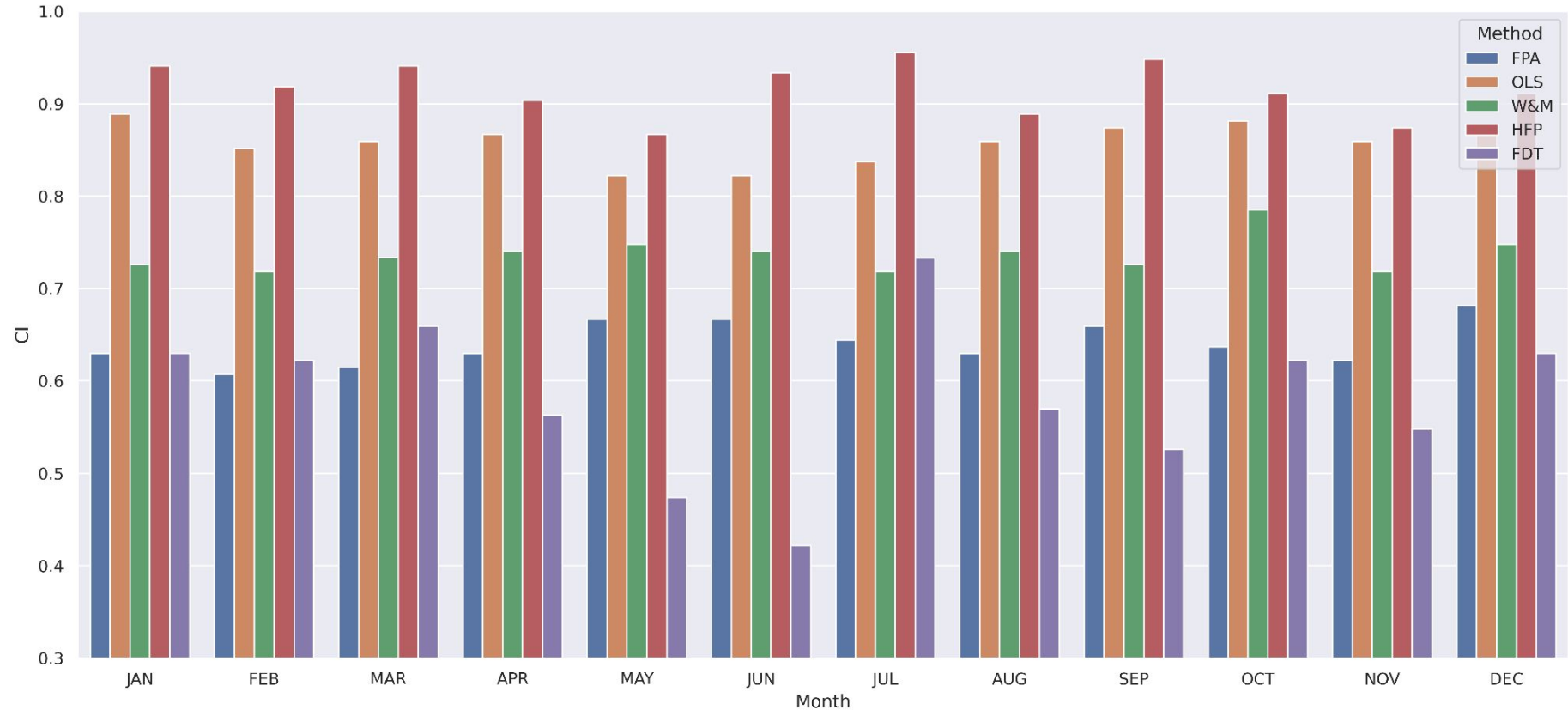
Risultati ottenuti
per il dataset sui
consumi energetici

Risultati in termini di Performance Index



- Le performance migliori, basate sulla metrica nota come *Root Mean Square Error*, sono ottenute dagli algoritmi OLS e HFP.
- Analizzando con attenzione i rispettivi sistemi generati da entrambi gli algoritmi, si evidenzia come:
 - l'algoritmo OLS ha un numero di MF per ciascuna variabile pari a 4;
 - al contrario, l'algoritmo HFP, soprattutto grazie al particolare meccanismo di selezione, ha un numero medio di MF utilizzato per ciascuna variabile pari a 2.5
 - Le performance dell'algoritmo Fuzzy decision tree risultano di poco inferiori, ad eccezione dei mesi estivi, dove le prestazioni ottenute sono molto simili a quelle ottenute con OLS e HFP

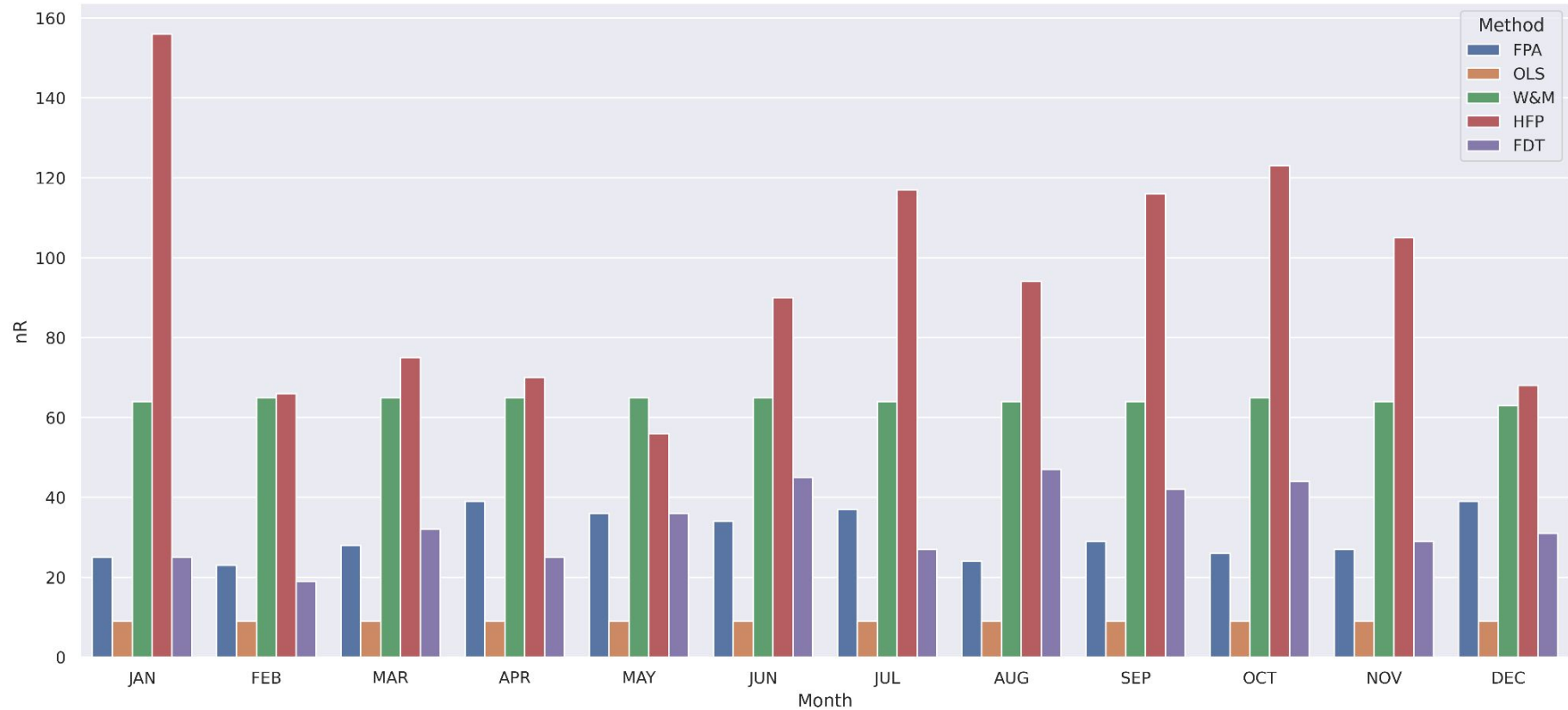
Risultati in termini di Coverage Index



Risultati in termini di Coverage Index

- L'indicatore di Coverage index rappresenta in termini percentuali gli esempi coperti dalle regole indotte.
- Un valore di coverage index elevato indica una maggiore copertura degli esempi attivi
- Tuttavia, una copertura troppo elevata potrebbe portare all'ottenimento di un sistema che tende a sovradattarsi ai dati, anziché generalizzare
- Valori di coverage troppo contenuti possono intaccare anche le prestazioni del sistema

Risultati in termini di complessità (numero di regole)



Risultati in termini di complessità (numero di regole)

- Oltre alle performance, un indicatore importante è la complessità
 - La complessità è data dal numero di regole indotte
 - Un numero troppo elevato di regole può portare ad ottenere un sistema poco efficiente
 - Il numero di regole deve essere mantenuto ragionevolmente piccolo
-
- Dai risultati ottenuti è possibile notare un numero elevato di regole indotte da parte del metodo HFP
 - Nonostante quest'ultimo abbia prodotto dei buoni risultati in termini di performance, la complessità rispetto all'utilizzo del metodo OLS risulta evidentemente più elevata

Conclusioni

- FisPro rappresenta un valido strumento per la costruzione di sistemi ad inferenza fuzzy, sia a partire da una base di conoscenza ma anche partendo dai dati
- L'interfaccia grafica gode di una buona usabilità, permettendo anche la visualizzazione dei dati in modalità 2D/3D
- Gli aggiornamenti del software non sono periodici
- L'esistenza di alcuni bug (alcuni già noti agli sviluppatori) può portare, in alcuni casi, ad interruzioni improvvise del sistema
- La documentazione non è del tutto allineata con le ultime versioni del software, ciò nonostante l'help da linea di comando può ovviare a questa problematica
- Il servizio di supporto garantisce assistenza in maniera celere, distribuendo snapshot con bugfix