

1- prima visualizzazione 01Auser.jpg  
2- primo taglio al 01/01/2006 user 02Auser

Commento: ci vengono forniti i dati di 5 pozzi, suddivisi in un gruppo a Nord, composto dai pozzi SAL, PAG, CoS e DIEC e da un pozzo a Sud, il pozzo LT2. Di queste 5 variabili, che esprimono il livello di profondità dell'acqua, in metri, nei pozzi, le nostre variabili target sono i livelli dei pozzi SAL e CoS a nord e LT2 a sud.

Tabella dei missing

A tibble: 27 x 3

variable	n_miss	pct_miss
<chr>	<int>	<dbl>
1 Depth_to_Groundwater_DIEC	4884	59.9
2 Depth_to_Groundwater_PAG	4347	53.3
3 Depth_to_Groundwater_CoS	3839	47.1
4 Depth_to_Groundwater_SAL	3609	44.3
5 Depth_to_Groundwater_LT2	3352	41.1
6 Rainfall_Piaggione	3224	39.5
7 Rainfall_Monte_Serra	2865	35.1
8 Rainfall_Gallicano	2859	35.1
9 Rainfall_Pontetetto	2859	35.1
10 Rainfall_Orentano	2859	35.1

# ... with 17 more rows

3- grafico 04AUSER\_filtered\_depth\_to\_groundwater andamento dei pozzi con presenti i missing + Grafico 05AUSER\_target\_depth\_to\_groundwater dettaglio dei pozzi target SAL, CoS e LT2 (degl  
tu se inserire entrambi i grafici o solo il secondo)

La parte verde e' il pozzo a sud LT2. I valori di profonditi' sono espressi in valore assoluto

4-grafico 12AUSER\_corr\_temp Temperatur + variabili target, tre pozzi

Non sembra che ci sia correlazione (la parte verde LT2 rappresenta l'unico pozzo a sud)

5- grafico 11AUSER\_corr\_rain + variabili target, tre pozzi, la parte verde e' il pozzo a sud LT2

6- grafico 14AUSER\_pozzi\_missing visualizzo la d'istituzione dei valori missing dei 5 pozzi, sia per i tre target e anche per gli altri pozzi. Studio come riempire i valori imputandoli con una semplice linear interpolation oppure utilizzando a funzione imputeTS dedicata alle serie temporali.

7- ecco il risultato con il grafico 05AUSER\_pozzi\_NO\_missing con i valori riempiti con la funzione input TS, per i pozzi (sia target, sia per gli altri pozzi).

8-outliers grafico 16box\_auser\_pozzi

+ i grafici 17, 18, 19 delle variabili target (cos, lt2, Sal). Dove possibile, porto i valori outliers al quartile ppiu' vicino, per esempio se ipotizzo un errore del sensore di misurazione. Per valori outliers ripetuti, come per esempio per il pozzo SAL, non e' possibile intervenire.

Per esempio, il pozzo CoS ha 4 valori di outliers pari a zero e decido di intervenire:

Grubbs test for one outlier:

data: auser8\$impCoS

G = 4.77370, U = 0.99569, p-value = 0.004673

alternative hypothesis: lowest value 0 is an outlie

Il risultato e' questo grafico 20auser\_boxplot (decidi tu se inserirlo oppure no)

9- grafico 23auser\_temp no outliers temp

Ma sul grafico img/auser/21auser\_temp.jpg noto che il sensore di pioggia di Ponte a Moriano, dal 2017 , riporta il valore 0 C consecutivo per mesi interi , sicuramente e' un valore errato e dal modello dovro' escludere questa variabile.

10-grafico 24auser\_rain.jpg there's a long tail si vedono alcuni outliers + grafico 26auser\_rain\_all visone d'insieme delle variabili pioggia (outliers)

11- checking out for missing - rain and temp

12- grafico 29auser\_rain\_missing2 vedo i missing per la pioggia e riempio con i dati di rbmete3 (Per le temperature non ci sono missing, non serve il grafico, specifico quanto scritto sopra che per il modello devo escludere la temperatura di Ponte a Moriano, che per mesi consecutivi porta il valore a zero)

Se vuoi riportare il codice per Monte Serra:

## manipulating temp data

```
rf_monteserra <- rf_monteserra_ms %>%
  dplyr::rename(date1 = filename) %>%
  mutate(date1 = gsub("./data/meteoMonteSerra/", "", date1),
         date1 = gsub(".csv", "", date1),
         date1 = gsub("([a-z])([[:digit:]])", "\\1 \\2", date1, perl = T)) %>%
  separate(date, into = c("weekday", "day")) %>%
  dplyr::select(-weekday) %>%
  unite(date_final, day, date1, sep = " ") %>%
  mutate(date_final = str_replace(date_final, "ago", "08"),
         date_final = str_replace(date_final, "gen", "01"),
         date_final = str_replace(date_final, "feb", "02"),
         date_final = str_replace(date_final, "mar", "03"),
         date_final = str_replace(date_final, "apr", "04"),
         date_final = str_replace(date_final, "mag", "05"),
         date_final = str_replace(date_final, "giu", "06"),
         date_final = str_replace(date_final, "lug", "07"),
         date_final = str_replace(date_final, "sett", "09"),
         date_final = str_replace(date_final, "ott", "10"),
         date_final = str_replace(date_final, "nov", "11"),
         date_final = str_replace(date_final, "dec", "12"),
         date_final = gsub(" ", "/", date_final),
         date_final = dmy(date_final)) %>%
  dplyr::rename(Date = date_final) %>%
  dplyr::select(Date, prec) %>%
  arrange(Date)
```

13-Missing Hydrometry grafico 30auser\_Hydrometry

I valori mancanti sono pochi si possono interpolare

14- grafico 31auser\_NO\_missing user senza I missing

15- abbiamo aggiunto la variabile stagione e la variabile neve

16-grafico 32auser\_target\_no\_missing per visualizzare andamento delle variabili da predire

17- grafico 33auser\_correlazione matrice di correlazione di tutte le variabili - la situazione si presenta complessa, le piogge sono abbastanza correlate tra loro e anche le temperature.

18- grafico 34auser\_temp.jpg temperature dove si nota l'errore del sensore di Monte a Moriano a zero da giugno 2017

19- grafico 35auser\_volumi

I valori di queste variabili

#sono negativi in quanto riguardano la quantità di acqua

#prelevata dal luogo indicato.

# si puo' pensare di di escludere le variabili che descrivono i

#serbatoi CC1 e CSA per la loro correlazione con altre variabili, e di scegliere le variabili CC2 e CSAL per la maggiore varianza

20- grafico 36auser\_H commento

##commento: osservo piaggione  
#nel modello si puo' pensare di rimuovere questa variabile anche perche'  
#molto correlata con altre, guardano la matrice di correlazione

## 21- grafico 37auser\_season\_target distribuzione delle variabili target per stagione

Commento commento

# per il pozzo LT2 non ci sono molte differenze per stagioni  
# per i pozzi CoS e per Sal i pozzi sono piu' profondi in estate e autunno  
# ricordo che abbiamo preso il valore assoluto (i valori sono negativi)

## 22- MODELLI

#####Random forest

-Target impCOS. Grafico 40auser\_RF\_cos.jpg

#Il pozzo PAG ha maggiore influenza sul modello.

#Ha un impatto sul modello oltre 3 volte maggiore rispetto alla

#seconda variabile più importante, che è il volume POL

#Le piogge dalle regioni incluse, si sono rivelate avere

#il minor impatto sul modello.

#### errore RMSE Test: 0.22 COS ####

Sample size of test (predict) data: 1106

Number of grow trees: 500

Average no. of grow terminal nodes: 441.69

Total no. of grow variables: 15

Resampling used to grow trees: swr

Resample size used to grow trees: 1106

Analysis: RF-R

Family: regr

% variance explained: 97.39

Test set error rate: 0.05

-Target impLT2

Grafico 41auser\_RF\_LT2

#Il volume CSAL ha maggiore influenza sul modello, le piogge hanno poca influenza

#### errore RMSE Test: 0.07 LT2 ####

Il volume CSAL ha maggiore influenza sul modello.

#Ha un impatto sul modello oltre 3 volte maggiore rispetto alla

#seconda variabile più importante, che è il volume CC2.

#

#Le piogge dalle regioni incluse, si sono rivelate avere

#il minor impatto sul modello.

Sample size of test (predict) data: 1106

Number of grow trees: 500

Average no. of grow terminal nodes: 439.236

Total no. of grow variables: 19

Resampling used to grow trees: swr

Resample size used to grow trees: 1106

Analysis: RF-R

Family: regr

% variance explained: 98.31

Test set error rate: 0

-Target pozzo impSAL

GRAFICO 42auser\_RF\_SAL

#### errore RMSE Test: 0.1 SAL ####

#Il pozzo DIEC ha maggiore influenza sul modello.

```
#Ha un impatto sul modello oltre 3 volte maggiore rispetto alla
#seconda variabile più importante, che è il pozzo PAG.
#
#Le piogge dalle regioni incluse, si sono rivelate avere
#il minor impatto sul modello.
```

### 23- MODELLI GBM

Avendo tre variabili target, ho creato tre dataset

pozzo\_SAL , pozzo\_LT2 , pozzo\_CoS

-POZZO COS

Per fare un esempio, riporto i dati di pozzo CoS

```
coef(pozzo_CoS_sw$finalModel, 10)
```

```
(Intercept)      Rainfall_Monte_Serra
-6.722509e-01      4.213340e-03
Temperature_Monte_Serra      Volume_POL
 1.135433e-02      -2.792359e-04
Volume_CC2      Volume_CSAL
-1.948930e-05      -1.120896e-05
Hydrometry_Monte_S_Quirico      imp2
 1.135015e+00      4.560863e-01
      imp4      SeasonAutumn
 1.083754e+00      6.755132e-01
Season.Spring
-3.611277e-01
```

```
#### pozzo cos GBM #### RMSE 0.2248 pozzo CoS GB ####
```

```
# A tibble: 6 x 2
```

```
var      rel.inf
<chr>    <dbl>
1 pozzo_PAG      59.3
2 Volume_POL     14.8
3 pozzo_DIEC     10.2
4 Season.Autumn   5.91
5 Volume_CC2      3.55
6 Hydrometry_Monte_S_Quirico  1.85
```

Grafici 43 e 44 — — —

÷÷÷//??Valori attuali e predetti grafico 45pozzo\_cos\_pred??? Mi sembra errato non mettere

- POZZO pozzo 3 LT2

```
#### RMSE 0.0865 pozzo LT2 GB ####
```

-

Grafico 46pozzo\_LT2\_GB. - quanto incidono le variabili:

```
# A tibble: 6 x 2
```

```
var      rel.inf
<chr>    <dbl>
1 Volume_CSAL  58.3
2 pozzo_DIEC   18.5
3 Volume_CC2    8.96
4 Volume_POL    6.32
5 Season.Spring  1.90
6 pozzo_PAG     1.83
```

Grafico 47auser\_pozzolt2\_features

Grafico 48pozzo\_lt2\_pred valori attuali e predetti??? Mi sembra errato non mettere

- POZZO SAL
- Grafico 49pozzo\_SAL\_GB
- ##### RMSE 0.1139 pozzo SAL GB #####
- INFLUENZA DELLE VARIABILI
- # A tibble: 6 x 2
- var rel.inf
- <chr> <dbl>
- 1 pozzo\_PAG 43.6
- 2 Hydrometry\_Monte\_S\_Quirico 14.5
- 3 pozzo\_DIEC 12.4
- 4 Volume\_POL 10.6
- 5 Volume\_CSAL 8.55
- 6 Volume\_CC2 3.28

Grafico 50auser\_pozzosal\_features

Grafico 51pozzo\_sal\_pred ??? Mi sembra errato non mettere