

# weather\_anomaly

Rodrigo Casella

n° matricola 599523

[r.casella2@studenti.unipi.it](mailto:r.casella2@studenti.unipi.it)

## Introduzione

*weather\_anomaly* è un piccolo tool che permette di individuare e visualizzare anomalie all'interno di serie temporali che rappresentano la temperatura media giornaliera. Un'anomalia è definita come l'osservazione di un dato che si discosta dalla predizione calcolata e che quindi cade al di fuori degli intervalli di fiducia della serie temporale. Tutte le osservazioni, le predizioni e bande di confidenza vengono salvate in un archivio rrd e presentate con un grafico.

I file letti devono essere in formato csv e seguire il seguente formato: il primo valore è il codice della stazione meteo, il secondo è la data della misurazione nel formato "YYYY-MM-DD" ed infine la temperatura al momento della misurazione in Celsius; altre informazioni sul formato sono reperibili al seguente indirizzo:

[https://www.ncei.noaa.gov/pub/data/cdo/documentation/GHCND\\_documentation.pdf](https://www.ncei.noaa.gov/pub/data/cdo/documentation/GHCND_documentation.pdf)

Le predizioni e le bande di confidenza sono calcolate usando l'algoritmo Holt-Winters implementato nella libreria di nDPI <https://github.com/ntop/nDPI>

## Prerequisiti

Per eseguire *weather\_anomaly* basta installare la libreria nDPI e tutti i suoi prerequisiti.

Successivamente nel Makefile bisogna impostare il proprio path della libreria nDPI.

## Esecuzione

Per compilare il codice basta eseguire `make all`.

L'esecuzione di *weather\_anomaly* richiede in modo obbligatorio un file csv che segua il formato precedentemente illustrato e la durata (in giorni) della stagionalità della serie temporale.

```
./bin/ weather_anomaly -f <percorso-file-csv> -s <periodo-stagione>
```

E' possibile impostare anche altri tipi di parametri, specialmente per l'algoritmo Holt-Winters:

-a <alpha> Parametro alpha per Holt-Winters

-b <beta> Parametro beta per Holt-Winters

-g <gamma> Parametro gamma per Holt-Winters

-r <ro> Parametro ro per Holt-Winters

-d <nome\_archivio> Imposta nome per archivio rrd

-i <nome\_immagine> Imposta nome immagine prodotto da rrdtool

-v Attiva la verbose mode (Vengono stampati sullo standard output anche i valori non anomali)

## Holt-Winters

Le predizioni sulla serie temporale vengono calcolate attraverso l'algoritmo Holt-Winters. Consiste in uno smoothing della serie temporale osservata tenendo in conto tre diverse componenti:

$$\text{level: } \ell_t = ((\alpha * y_t) / s_{t-1}) + (1 - \alpha) * (\ell_{t-1} + b_{t-1})$$

$$\text{trend: } b_t = \beta * (\ell_t - \ell_{t-1}) + (1 - \beta) * b_{t-1}$$

$$\text{componente stagionale: } s_x = ((\gamma * y_t) / \ell_t) + (1 - \gamma) * s_{t-1}$$

Ci sono due diversi tipi di stagionalità: additiva e moltiplicativa. Nel nostro caso la variazione di temperatura ha una stagionalità di tipo additivo poiché è generalmente costante. Quindi la nostra predizione per l'osservazione al tempo  $t$  sarà uguale a:

$$y^{\wedge}_t = \ell_{t-1} + b_{t-1} + s_{t-1} \% L$$

## RRDTool

Le temperature osservate vengo salvate con cadenza giornaliera in archivi rrd.

La struttura dell'archivio rrd è la seguente:

DS:data:GAUGE:2d:-273:5000

RRA:AVERAGE:0.5:1:period

RRA:HWPREDICT:rows:alpha:beta:season\_period

Quindi abbiamo una fonte dati che raccoglie la temperatura osservata ogni giorno e due archivi round-robin: uno raccoglie le osservazioni durante tutto il periodo osservato, mentre il secondo calcola e raccoglie le predizioni usando l'implementazione interna di rrdtool dell'algoritmo Holt-Winters.

Vengono inoltre modificati i seguenti parametri:

La lunghezza dell'archivio FAILURES creato implicitamente da HWPREDICT, il valore di ro per il calcolo delle bande di confidenza, il valore di gamma per Holt-Winters e la finestra ed il numero di violazioni per catalogare un punto come anomalo.

Il grafico prodotto rappresenterà la serie temporale delle temperature osservate, le predizioni e le bande di confidenza, inoltre con delle linee verticali verranno segnalate le anomalie in corrispondenza del giorno dove sono state riscontrate.

## Test

Per poter testare il programma sono presenti tre file csv nella cartella *test*.

Una possibile esecuzione potrebbe essere:

```
./bin/weather_anomaly -f test/01012019-31122021.csv -s 365
```

Output:

When	Value	Prediction	Lower	Upper	Out	[Band]
01/Jan/2020 00:00:00	1.700	7.821	6.675	8.967	ANOMALY	[1.146]
02/Jan/2020 00:00:00	-3.900	0.772	-0.662	2.207	ANOMALY	[1.434]
04/Jan/2020 00:00:00	6.100	7.769	6.307	9.230	ANOMALY	[1.461]
05/Jan/2020 00:00:00	2.200	15.643	12.777	18.510	ANOMALY	[2.867]
06/Jan/2020 00:00:00	0.000	8.009	4.800	11.218	ANOMALY	[3.209]
07/Jan/2020 00:00:00	-1.100	4.218	0.879	7.556	ANOMALY	[3.339]
08/Jan/2020 00:00:00	-1.700	10.785	6.763	14.806	ANOMALY	[4.021]
[...]						
13/Nov/2021 00:00:00	3.300	26.156	6.216	46.096	ANOMALY	[19.940]
07/Dec/2021 00:00:00	0.600	14.818	1.726	27.911	ANOMALY	[13.092]
08/Dec/2021 00:00:00	0.000	15.763	2.408	29.118	ANOMALY	[13.355]
13/Dec/2021 00:00:00	2.800	16.771	3.197	30.345	ANOMALY	[13.574]
18/Dec/2021 00:00:00	4.400	23.130	9.252	37.008	ANOMALY	[13.878]
19/Dec/2021 00:00:00	3.300	23.249	8.988	37.510	ANOMALY	[14.261]
20/Dec/2021 00:00:00	-6.700	20.168	5.203	35.132	ANOMALY	[14.965]
21/Dec/2021 00:00:00	-6.100	11.698	-3.528	26.924	ANOMALY	[15.226]

Grafico prodotto:

