**Descrizione del dataset**

Il dataset contiene i dati relativi all’inquinamento negli Stati Uniti raccolti da U.S EPA giornalmente dal 2000 al 2016; in particolare gli agenti presi in considerazione sono NO2(diossido di azoto), O3 (ozono), CO (monossido di carbonio), SO2 (BIOSSIDO DI ZOLFO).

I campi sono:

State Code 🡪 Codice identificativo univoco per Stato

County Code 🡪 Codice identificativo univoco per Contea

Site Num 🡪 identificativo del sito di rilevazione

Address 🡪 indirizzo del sito di rilevazione

State 🡪 Stato del sito di rilevazione

County 🡪 Contea del sito di rilevazione

City 🡪 Città del sito di rilevazione

Date Local 🡪 data locale della rilevazione

NO2 Units 🡪 unità di misura di NO2

NO2 mean 🡪 valor medio di NO2 rilevato

NO2 first\_max\_value 🡪 valore massimo di NO2 rilevato

NO2 first\_max\_hour 🡪 ora di rilevazione del valore massimo di NO2  relativo al giorno considerato

NO2 AQI 🡪 Air Quality Index, indice indicativo della qualità dell’aria relativo alla concentrazione di NO2 nell’aria

O3Units 🡪 unità di misura di O3

O3 mean🡪 valor medio di O3 rilevato

O3 first\_max\_value 🡪 valore massimo di O3 rilevato

O3 first\_max\_hour 🡪 ora di rilevazione del valore massimo di O3 relativo al giorno considerato

O3 AQI 🡪Air Quality Index, indice indicativo della qualità dell’aria relativo alla concentrazione di O3 nell’aria

SO2 Units 🡪 unità di misura di SO2

SO2 mean 🡪 valor medio di SO2 rilevato

SO2 first\_max\_value 🡪 valore massimo di SO2 rilevato

SO2 first\_max\_hour 🡪 ora di rilevazione del valore massimo di SO2 relativo al giorno considerato

SO2 AQI 🡪 Air Quality Index, indice indicativo della qualità dell’aria relativo alla concentrazione di SO2 nell’aria

CO Units 🡪 unità di misura di CO

CO mean 🡪 valor medio di COrilevato

CO first\_max\_value 🡪 valore massimo di COrilevato

CO first\_max\_hour 🡪 ora di rilevazione del valore massimo di COrelativo al giorno considerato

CO AQI 🡪 Air Quality Index, indice indicativo della qualità dell’aria relativo alla concentrazione di CO nell’aria

**Funzionalità**

L’obiettivo del progetto è quello di effettuare una predizione dei valori degli agenti inquinanti (campi evidenziati in giallo), sulla base dei dati relativi alla posizione geografica (campi evidenziati in verde).

Per eseguire la prediction attraverso l’utilizzo di algoritmi di Machine Learning, utilizzeremo esclusivamente i campi numerici.

Per il training verranno utilizzati i dati relativi agli anni dal 2010 al 2014 per poi effettuare una predizione sugli anni 2015-16, in modo da permettere un confronto oggettivo con i dati presenti nel dataset relativi a quel periodo.

La motivazione della scelta degli anni considerati risiede nel fatto che le singole stazioni di rilevazione non presentano dati pienamente consistenti e/o dalla quantità soddisfacente negli anni precedenti al 2010.