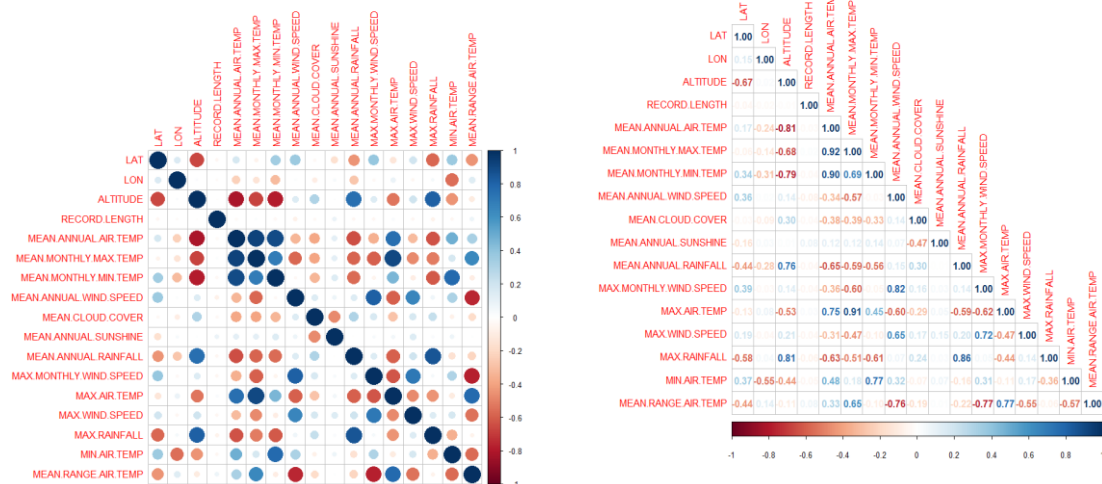


Regresja wieloraka liniowa. Trenowanie i ewaluacja modeli – część 1

1. Wczytaj zestaw danych '**weather.csv**'. Dane zawierają informacje na temat 599 stacji pogodowych opisanych przez 22 cechy/zmienne. Zestaw danych został pobrany z niemieckiego serwisu pogodowego (www.dwd.de/EN/Home). Przy wczytywaniu użyj argumentu `stringsAsFactors = FALSE`.
2. Sprawdź typy wszystkich zmiennych.
3. W kolejnych ćwiczeniach będziemy tworzyć model predykcji średnich rocznych opadów. Zdecyduj, które zmienne nie wniosą żadnych istotnych informacji do naszego modelu. Wyklucz je ze zbioru danych.
4. Usuń wszystkie wiersze w których znajduje się jedna lub więcej brakująca wartość. Podpowiedź: użyj funkcji **`complete.cases()`**
5. Sprawdź wymiar nowego zestawu danych (po odrzuceniu wierszu z brakującymi wartościami oraz wykluczonych zmiennych).
6. Podziel utworzony zbiór danych na zbiór uczący i testowy w sposób losowy. W tym celu zaimportuj bibliotekę **`caTools`** i użyj funkcji **`sample.split()`**. Podziel dane w taki sposób, aby 80% znajdowało się w zbiorze treningowym, a 20% w zbiorze testowym. Utwórz dwa odrębne zestawy danych – dla zbioru treningowego i testowego.
7. Zmienną objaśnianą modelu będzie średnia roczna wielkość opadów (MEAN.ANNUAL.RAINFALL). Wyodrębnij tą zmienną w postaci wektora dla zbioru uczącego i testowego. Utwórz macierz dla obydwu zbiorów składające się ze zmiennych objaśniających.
8. Przedstaw macierz wykresów rozrzutu parami używając funkcji **`ggpairs()`** z pakietu **`Ggally`**.
9. Oblicz korelację pomiędzy parami wszystkich zmiennych.
10. Korzystając z funkcji **`corrplot()`** z pakietu **`corrplot`** przedstaw wartości korelacji pomiędzy kolejnymi zmiennymi. Domyślnie funkcja ta wyświetli korelacje używając kolorowych okręgów. Zdefiniuj argumenty funkcji `corrplot()` tak aby przedstawić wartości korelacji w sposób liczbowy. Ponieważ macierz korelacji jest macierzą symetryczną można dodatkowo ustawić argument aby wyświetlić macierz trójkątną górną lub dolną.



11. Do dalszej analizy włączymy tylko te kombinacje par zmiennych, których wartość bezwzględna współczynnika korelacji jest większa lub równa 0.5.
12. Przedstaw macierz wykresów rozrzutu parami używając funkcji **ggpairs()** dla zmiennych, które pozostały wybrane do dalszej analizy.

