

# ZZN projekt – 1. časť

## Databáze počasí v Austrálii

Natália Ivanisková, xivanin00  
Emma Krompaščíková, xkromp00

## 1 Zadanie

### 1.1 Popis dát

CSV súbor obsahuje dáta o počasí v Austrálii a má 145 460 záznamov. Dáta boli zaznamenané na 49 miestach, avšak pre každé miesto sa časové rozmedzie záznamov líši. Maximálne časové rozpätie záznamov je takmer 10 rokov od 01.11.2007 do 25.06.2017. Každý záznam je popísaný atribútmi typu string alebo float:

- **[string]:** Date, Location, WindGustDir, WindDir9am, WindDir3pm, RainToday a RainTomorrow
- **[float]:** MinTemp, MaxTemp, Rainfall, Evaporation, Sunshine, WindGustSpeed, WindSpeed9am, WindSpeed3pm, Humidity9am, Humidity3pm, Pressure9am, Pressure3pm, Cloud9am, Cloud3pm, Temp9am a Temp3pm

Hodnoty *Evaporation*, *Sunshine*, *Cloud9am* a *Cloud3pm* vo viac ako 38% záznamov chýbajú, a preto tieto atribúty budú vylúčené. V prípade ostatných atribútov chyba menej ako 11% hodnôt a to bude ošetrené podľa potrieb konkrétnej dolovacej úlohy.

### 1.2 Návrh dolovacích úloh

#### 1.2.1 Porovnanie teplôt medzi rokmi pre jednotlivé mesiace

Pre každý mesiac budú spočítané priemerné, minimálne a maximálne denné teploty. Teploty pre jednotlivé mesiace budú porovnané medziročne a dá sa skúmať, či sa teploty zvyšujú alebo znižujú. Toto porovnanie je potrebné uskutočniť pre každé miesto osobitne.

#### 1.2.2 Klasifikácia – prší alebo neprší?

Na základe nameraných hodnôt tlaku, vlhkosti a oblačnosti je možné klasifikovať, či bude v daný deň pršať. V datasete sa nachádza atribút *Rainfall*, ktorý popisuje množstvo zrážok v *mm*, a keď je jeho hodnota väčšia ako 1, tak atribút *RainToday* je *Yes*. Preto atribút *Rainfall* odstránime, a atribút *RainToday* bude použitý ako label pre klasifikáciu do dvoch tried *Yes* alebo *No*.

### **1.2.3 Asociačné pravidlá pre extrémne počasie**

Na základe údajov o teplote, vlhkosti, sile a smere vetra, tlaku a oblačnosti by sa dalo predpovedať extrémne počasie vo forme asociačných pravidiel. Pravidlá sa môžu kvôli geografickej polohe miest líšiť, preto by bolo vhodné pridať atribút popisujúci polohu alebo podnebné pásmo, v ktorom sa miesto v rámci Austrálie nachádza. V rovnakom podnebnom pásme by mali byť asociačné pravidlá podobné.