

1. Pomocí metody PRIM zjistěte, zda je rozdíl v morfologických parametrech (hodnoty jsou v mm) mezi samečkami a samičkami u želv (samečci = 1, samičky = 2).

2. Studie různých typů klasifikací skla byla motivována vyšetřováním trestných činů. Datový soubor obsahuje 214 vzorků šesti druhů různého skla a devět prediktorů – různé chemické parametry skla.

typy skla: (class)

A okenní sklo (z domu)1

B okenní sklo (z domu)2

C okenní sklo (z auta)1

D okenní sklo (z auta)2

E obalové sklo

F sklo z nádobí

G reflektorové sklo

Prediktory: index lomu (RI), Na, Mg, Al, Si, K, Ca, Ba, Fe

Pomocí metody Random Forest vytvořte optimální les a zhodnoťte výsledky. Který parametr je nejvýznamnější, které skupiny odlišuje a pro které jeho hodnoty jsou kategorie odlišitelné. Zhodnoťte překryv kategorií. Kolik vzorků reflektorového skla bylo zařazeno do správné kategorie s pravděpodobností  $> 0.5$ ?

3. Pomocí metody CART zjistěte, zda lze odlišit dva regiony na základě uvedených parametrů prostředí a pokryvností rostlin. Mají větší vliv klimatické parametry nebo zastoupení mechů a bylin?

4. Zjistěte metodou MARS, zda koncentrace polyaromatických uhlovodíků (PAHs) v půdě závisí na uvedených parametrech. PAHs vznikají spalováním v továrnách, domácích topeništích nebo spalovacích motorech dopravních prostředků. Za jakých podmínek je koncentrace nejvyšší? Které parametry jsou významné? Pomocí vytvořeného modelu předpovězte hodnoty koncentrace pro podmínky v listě koncentrace PAH\_predikce.

(parametry: corg = procento organického uhlíku v půdě, nad\_vyska = nadmořská výška, vzd\_sidlo = vzdálenost měřené lokality od zastavěných oblastí (vesnice, města...) (m), vzd\_prum = vzdálenost měřené lokality od průmyslových oblastí (m), ROCTEP = průměrné roční teploty, SRAZKY = průměrné roční srážky, vzd\_silnice = vzdálenost od silnice, PM10 = objem prachových částic v ovzduší, hustota\_obyv = počet obyvatel, riziko = vyšší hodnoty znamenají větší riziko přítomnosti starých zátěží (bývalé letiště, skládky, lomy..)