

Data:

Parkinsonova choroba – analýza mluvené řeči

Na následující adrese je možné nalézt podrobný popis dat včetně podrobného popisu experimentů.

http://sami.fel.cvut.cz/Articles/Rusz_et_al_Acoustic_analysis_of_Parkinsons_disease.pdf

Popis dat:

Parkinsonova choroba je chronické neuro-degenerativní onemocnění. Jedním z projevů choroby je i špatné mluvení. Otázkou je zda na základě analýzy hlasu jsme schopni diagnostikovat toto závažné onemocnění.

Atributy:**Vstup**

1. Pořadové číslo
- 2.
3. Pohlaví
4. Věk
5. jitter(local)
6. jitter(rap)
7. jitter(ppq5)
8. jitter(ddp)
9. shimmer(local)
10. shimmer(local, dB)
11. shimmer(apq3)
12. shimmer(apq5)
13. shimmer(apq11)
14. shimmer(dda)
15. NHR
16. HNR
17. DFA
18. RPDE
19. PPE
20. RFPC
21. SDCV
22. SPLD
23. F0 SD

Výstup:

2. Subject : PN – Parkinsonova nemoc, KS – Kontrolní skupina

Úkoly: Každému studentovi je přidělena celá datová sada**Předzpracování dat:**

- 1) Nalezněte průměr, medián, minimum, maximum a standardní odchylku pro tři po sobě jdoucí atributy (například 1-3, nebo 8-10, apod.)
- 2) Zkonstruujte pro tyto atributy histogramy. Můžete na základě histogramů nalézt nějaké odlehle hodnoty.
- 3) Existují mezi těmito atributy korelace?

Strojové učení:

- 1) Vyberte dvě metody strojového učení a pokuste se klasifikovat pacienty na trpící chorobou a kontrolní skupinu.
- 2) Zkonstruujte křivku učení pro obě vybrané metody a jejich výsledky porovnejte.

- 3) Zkuste další metodu validace dat.
- 4) Který atribut je pro klasifikaci nejdůležitější.

Shluková analýza:

- 1) Existují v datech významné podskupiny?
- 2) Charakterizujte alespoň jednu takovou podskupinu.