Situacija brojevi-klase je ovakva (sve kvantitete pišem u odnosu na filtrirani dataset nakon izbacivanja nepotpunih instanci, dakle neće se podudarati s excel tablicom, ali proporcije otprilike hoće):

* HISTOLOGY (u diplomskom **HD**)
  + Koristimo isključivo oznake koje smo dobili iz Graza:
    - No – zakodirano kao klasa 0 – 44 instance
    - Simple – zakodirano kao klasa 1 – 494 instance
    - Complicated – zakodirano kao klasa 2 – 158 instanci
* HISTOLOGY BINARY (u diplomskom **HBD**)
  + Koristimo isključivo oznake koje smo dobili iz Graza:
    - No – zakodirano kao klasa 0 – 44 instance
    - Yes – zakodirano kao klasa 1 – 652 instance
* POSTOP (u diplomskom **PD**)
  + Koristimo oznake kako smo se dogovorili kako bi smanjili broj klasa (grupirali srodne klase)
    - Simple – zakodirano kao klasa 0 – 558 instanci
    - Complex (complex, complex and abscess) – zakodirano kao klasa 1 – 98 instanci
    - Other (perforated, gangranos) – zakodirano kao klasa 2 – 39 instanci
* POSTOP BINARY (u diplomskom **PBD**)
  + Koristimo oznake kako smo se dogovorili kako bi dobili binarnu klasifikaciju
    - No – podrazumijeva instance s dijagnozom Simple (kakve su u PD) – zakodirano kao klasa 0 – 558 instanci
    - Yes – podrazumijeva sve ostale instance (complex, complex and abscess, perforated, gangranos) – zakodirano kao klasa 1 – 137 instanci
* NOVI DATASET POSTOP 1 (**NDP1**) – za isprobavanje predloženog da klasificiramo samo manje zastupljene klase zasebno:
  + Complex – zakodirano kao klasa 0 – 68 instanci
  + Complex and abscess – zakodirano kao klasa 1 – 30 instanci
  + Perforated, Gangranos – zakodirano kao klasa 2 – 39 instanci
  + AVG ACC na NN 5-foldu 0.595671
* NOVI DATASET POSTOP 2 (**NDP2**) – za isprobavanje predloženog da klasificiramo samo manje zastupljene klase zasebno (binarno):
  + Complex, Complex and abscess – zakodirano kao klasa 0 – 98 instanci
  + Perforated, Gangranos – zakodirano kao klasa 1 – 39 instanci
  + AVG ACC na NN 5-foldu 0.742857
* NOVI DATASET HISTOLOGY 1 (**NDH1**) – za isprobavanje predloženog da klasificiramo samo manje zastupljene klase zasebno (samo binarno budući da su samo tri klase ukupno):
  + No – zakodirano kao klasa 0 – 44 instance
  + Complicated – zakodirano kao klasa 1 – 158 instanci
  + AVG ACC na NN 5-foldu 0.888258
* NOVI DATASET HISTOLOGY 2 (**NDH2**) – za isprobavanje predloženog da klasificiramo samo najzastupljenihe klase (samo binarno budući da su samo tri klase ukupno):
  + Simple – zakodirano kao klasa 0 – 494 instance
  + Complicated – zakodirano kao klasa 1 – 158 instanci
  + AVG ACC na NN 5-foldu 0.848352

Rezultati vezani za dataset iz rada [Using Machine Learning to Predict the Diagnosis, Management and Severity of Pediatric Appendicitis](https://www.frontiersin.org/articles/10.3389/fped.2021.662183/full). Podaci i rad dostupni na [GitHub - i6092467/pediatric-appendicitis-ml: Using ML to predict the diagnosis, management, and severity of pediatric appendicitis](https://github.com/i6092467/pediatric-appendicitis-ml).

Dataset se u radu dijeli na tri zasebna skupa, slično kao što smo radili za diplomski – Diagnosis by Criteria (**DBC**), Management (**MGMT**) i Severity (**SEV**).

High-level pregled (test set accuracy s najboljim parametrima iz grid searcha prema average accuracy na 5-foldu (osim OneR)) (ostatak dostupan u **reports** folderu):

|  |  |  |  |
| --- | --- | --- | --- |
| Podaci  Rezultat | **DBC** | **MGMT** | **SEV** |
| Decision tree | 0.91 | 0.92 | 0.90 |
| OneR | 0.71 | **0.94** | 0.90 |
| Sequential covering | 0.91 | 0.86 | 0.84 |
| Random forest | **0.94** | **0.94** | **0.92** |
| MLP | 0.84 | 0.87 | **0.92** |

Svi ostali rezultati su u report folderu na github.com/ivancrg/app

**Objašnjenje prefiksa u folderu reports**

|  |  |
| --- | --- |
| **Prefiks** | **Model** |
| dt | Decision tree |
| oner | One Rule |
| seqcov | Sequential covering |
| rf | Random forest |
| nn | Neural network / MLP |

**Postfiks \_best\_params** – označava najbolje parametre dobivene kroz 5-fold CV

**Postfiks \_gs** – grid search korišten

**Postfiks \_classification\_report** – podaci o F1-scoreu, accuracy, recall, precision, support

**Postfiks \_confusion\_matrix** – confusion matrix

**Postfiks \_cv** – accuracy u CV

**Postfiks \_fimp** – feature importance plot