Универзитет у Крагујевцу

Факултет инжењерских наука



Семинарски рад из предмета:

ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА

Тема:

Абалоне

Студент: Професор:

Милош Николић 554/2015 др. Весна Ранковић, ред. проф.

Крагујевац 2023.

Садржај

[1. Поставка задатка 2](#_Toc136104372)

[2. Увод 3](#_Toc136104373)

[3. Припрема података 4](#_Toc136104374)

[3. 1. Учитавање података и елиминисање очигледних грешака при мерењу 4](#_Toc136104375)

[3. 2. Визуелизација и обрада података 7](#_Toc136104376)

[4. Неуронска мрежа 12](#_Toc136104377)

[4. 1. Подела скупа података 12](#_Toc136104378)

[4. 2. Избор најбољих параметара 12](#_Toc136104379)

[4. 3. Тестирање мреже 13](#_Toc136104380)

[5. Закључак 14](#_Toc136104381)

[6. Литература 14](#_Toc136104382)

# 1. Поставка задатка

Дата је поставка проблема и на:   
<https://archive.ics.uci.edu/ml/datasets/Abalone>

Представљени проблем треба решити применом неуронских мрежа, генетским алгоритмом или неком од комбинација алгоритама вештачке интелигенције. Као програмски језик се препоручују Јава, C++ и сл. Матлаб се може искључиво користити за брзу проверу изабраног алгоритма и као верификација добијених резултата.

У самом семинарском раду обавено је:   
- Објаснити задати проблем, улазне и излазне податке   
- Визуализовати податке   
- Објаснити коришћене алгоритме и оправдати њихово коришћење  
- Дискутовати резултате и извести закључке  
- Објаснити предности и мане коришћене методе

Семинарски рад је потребно послати на [tijanas@kg.ac.rs](mailto:tijanas@kg.ac.rs) (изворни код програма, извршна верзија, помоћне датотеке итд.) и предати у писаном облику најкасније до **28.05.2023.**

Писани део семинарског рада повезан у спиралу треба да садржи следеће делове:  
- насловну страна са јасно написаним основним подацима  
- поставку задатка  
- опис делова програма са илустрацијама и самим изворним кодом   
- списак коришћене литературе

Кандидат је у обавези да приликом одбране практичном демонстрацијом рада програма детаљно објасни рад појединих делова програма.

У Крагујевцу 28.03.2023. године

# 

# 2. Увод

Абалоне је велики морски пуж са изузетно богатим и укусним месом. Због своје цењености и велике потражње доведен је до ивице изумирања на Западној обали у Сједињеним Америчким Државама.

   
***Слика 1.*** *Абалоне (морски пуж)*

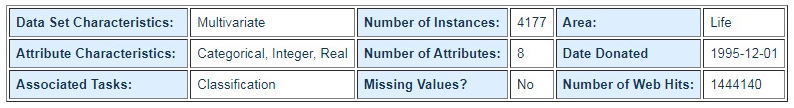
Једна од важнијих карактеристика која утиче на искористивост и цену абалонеа јесте његова старост. Да би се она одредила обично је потребно пребројавати број прстена под микроскопом, а затим на тај број додати 1,5. То је веома заморан посао који одузима доста времена. Међутим постоје мерења која су једноставнија за извршење, а могу помоћи да се старост абалонеа одреди са високом прецизношћу.

У овом задатку ће бити покушана обука неуронске мреже на тим подацима тако да буде способна да за нови, непознати тест примерак, одреди старост са што већом прецизношћу. Подаци су прикуљани мерењем више хиљада примерака.

За израду семинарског рада биће коришћен програмски језик *Python 3*, а због боље прегледности кода биће коришћен *Jupyter Notebook*.

# 3. Припрема података

Скуп улазних података се може преузети са сајта <https://archive.ics.uci.edu/ml/datasets/Abalone> при врху странице кликом на линк под називом *Data Folder.* Назив фајла јесте *Abalone.dat*. Овај фајл преводимо у датотеку са екстензијом *.csv*. У њему се не налазе називи атрибута, па их је потребно преузети кликом на линк *Data Set Description.* Када смо то урадили потребно је проучити скуп података (слика 2).

  
***Слика 2.*** *Генералне информације везане за скуп података*

Са приложене слике примећује се да постоји 8 атрибута и да су мешовитог типа (целобројни, реални, категорички), а укупан број инстанци је 4177. У овом скупу не постоје недостајуће вредности.

Атрибути су: пол, дужина, пречник, висина, укупна тежина, тежина меса, тежина утробе и тежина љуске. А вредност коју треба предвидети јесте број прстена.

## 3. 1. Учитавање података и елиминисање очигледних грешака при мерењу

Првобитно се увози библиотека *pandas* из које ће се позвати функција за учитавање *csv* датотеке, а затим се учитава сама датотека (слика 3). Вршимо приказ првих 5 редова како бисмо имали бољи увид у податке.

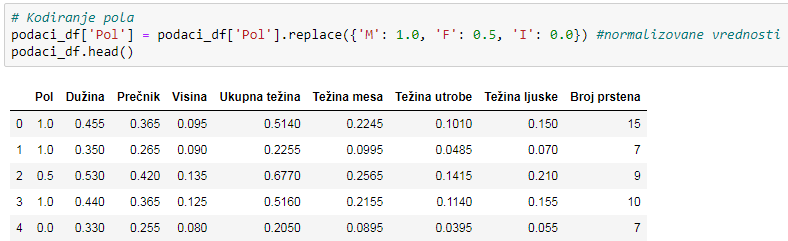
  
***Слика 3.*** *Учитавање података*

Када су подаци учитани потребно је проверити да ли смо успешно превели *.dat* датотеку и да ли подаци одговарају опису који је дат на сајту: <https://archive.ics.uci.edu/ml/datasets/Abalone>

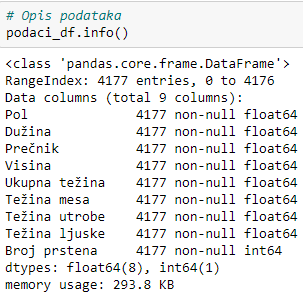
Из излаза овог дела примећујемо да је атрибут Пол словног типа (слика 4).

  
***Слика 4.*** *Излаз функције head()*

Кодираћемо вредности овог атрибута (код је приказан на слици 5) тако што ћемо заменити слово „M“ бројчаном вредношћу 1.0, слово „F“ бројем 1.5, а слово „I“ са 0.0. Ове бројеве смо одабрали јер примећујемо да сви остали подаци узимају вредности од 0 до 1.

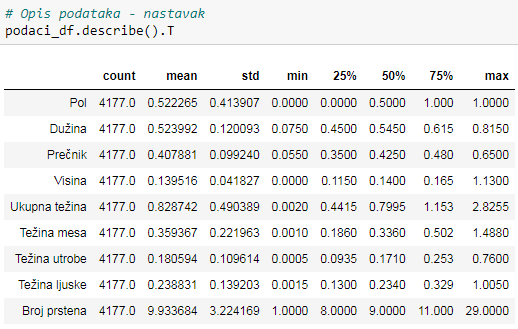
  
***Слика 5.*** *Кодирање атрибута Пол*

Да бисмо проверили остале информације користимо команду за опис података (слика 6).

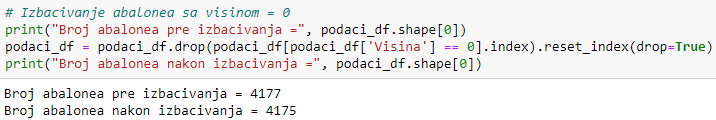
   
***Слика 6.*** *Опис података*

Из приложеног се види да постоји 9 колона, од којих ће једна колона бити за класификацију. Сви осим тог атрибута су реалног типа, док је тај атрибут целобројан. Број инстанци је 4177, а обележени су индексима од 0 до 4176. Такође видимо да нема недостајућих вредности.

За још детаљнији приказ користимо код приказан на слици испод (слика 7).

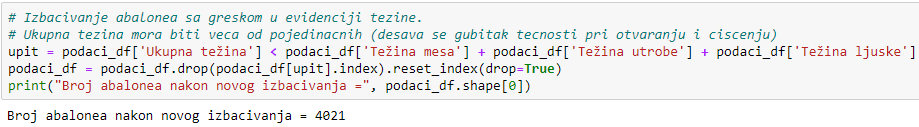
  
***Слика 7.*** *Детаљнији опис података*

Оно што је важно приметити јесте да постоје елементи чија је вредност висине једнака нули. Како је то немогуће испитаћемо који су то елементи и искључити их из даљег разматрања (слика 8). Након тога потребно је ресетовати индексирање.

   
***Слика 8.*** *Избацивање абалонеа са погрешном висином*

У наставку ћемо проверити да ли постоји грешка у евиденцији тежина. Укупна тежина абалонеа мора бити већа од тежина меса, утробе и љуске. Разлог зашто није једнака јесте губитак течности, односно крви при отварању и чишћењу абалонеа.

На слици испод (слика 9) видимо да се грешка десила у великом броју случајева (154).

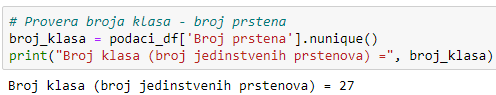
   
***Слика 9.*** *Избацивање абалонеа са погрешном евиденцијом тежина*

## 3. 2. Визуелизација и обрада података

Из хистограма испод (салика 10) видимо да је расподела података углавном нормална и да немамо вредности које драстично одскачу. Према томе у наставку ћемо сматрати да су подаци сада исправни и наставити са даљим током рада.

   
***Слика 10.*** *Хистограми за све атрибуте*

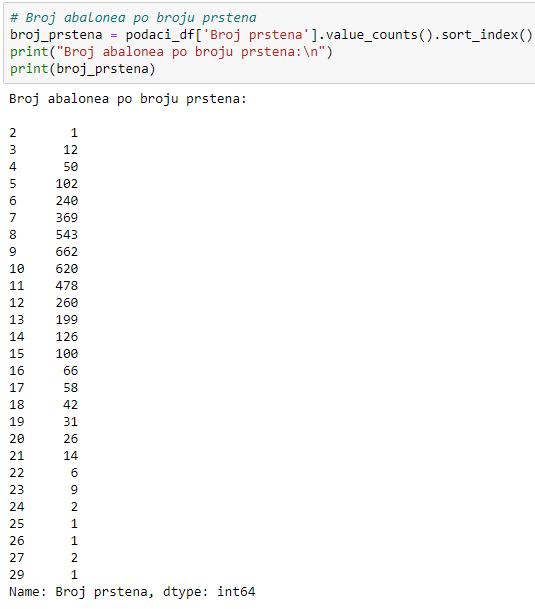
Проверићемо колико имамо јединствених класа (слика 11).

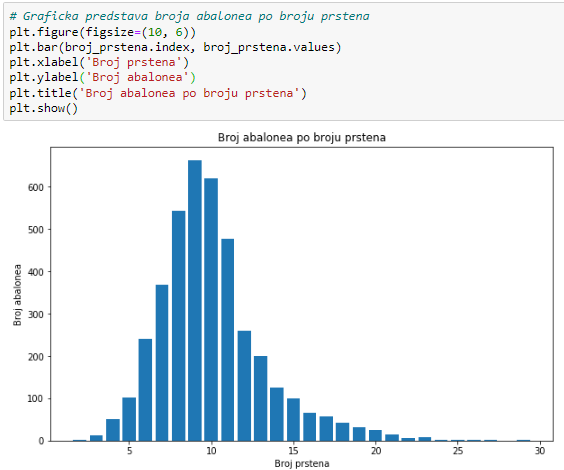
   
***Слика 11.*** *Број класа*

Видимо да је број класа велики, а бројчана разлика између њих мала, односно једнака јединици. У таквом случају било би веома тешко исправно класификовати те податке. Због тога је битно да смањимо број класа на логички исправан начин који неће утицати на примену у пракси.

Како се абалоне користи у исхрани идеја за поделу на класе која се намеће јесте подела на младе, зреле и старе. Да бисмо утврдили да ли је ова подела могућа и исправна потребно је извршити пар испитивања.

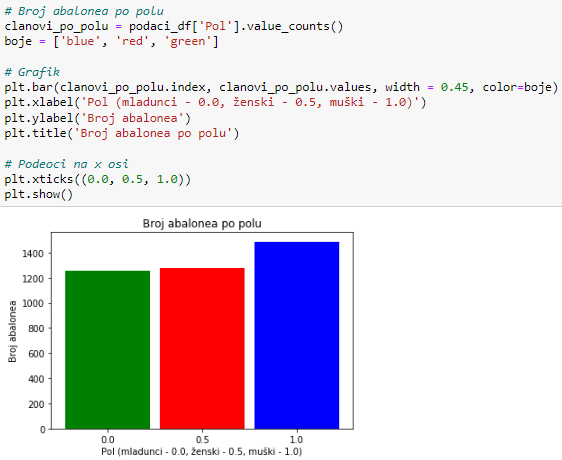
Са следеће две слике (слика 12 и слика 13) погледаћемо број абалонеа по броју прстена.

   
***Слика 12.*** *Број абалонеа по броју прстена*

   
***Слика 13.*** *График броја абалонеа по броју прстена*

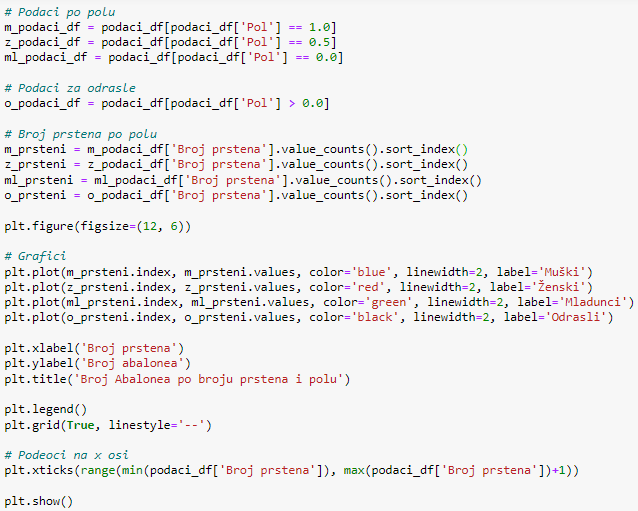
На основу двеју слика видимо да велика већина абалонеа има од 3 до 23 прстена. Остали број прстена има једна или две јединке. Према томе највише пажње при подели на категорије ћемо обраћати управо на овај део.

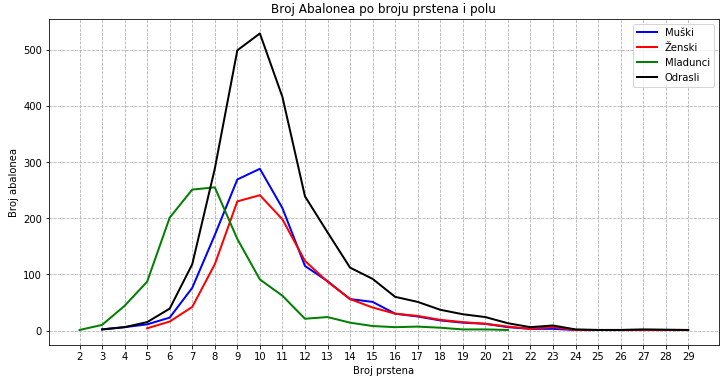
У наставку ћемо погледати број абалонеа по полу (слика 14).

   
***Слика 14.*** *График броја абалонеа по полу*

Видимо да је број јединки по полу сличан узимајући у обзир количину података. Како се при самом мерењу наглашава важност младунаца унутар пола, било би пожељно да пик њихових чланова припада категорији младих. Остатак мушких и женских јединки ћемо привремено спојити у још једну групу под називом одрасли.

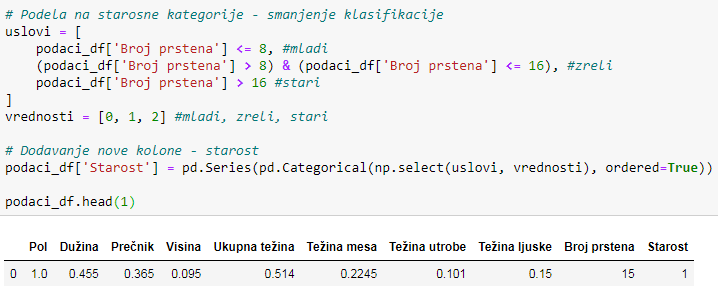
Сада ћемо приказати број абалонеа по броју прстена и полу (заједно са одраслима). Код (слика 15) и график (слика 16) су дати у наставку.

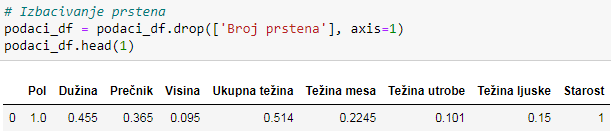
   
***Слика 15.*** *Код за приказ графика: број абалонеа по броју прстена и полу*

   
***Слика 16.*** *График: број абалонеа по броју прстена и полу*

Видимо да је облик графика за абалоне мушког и женског пола сличан, па можемо да кажемо да група „одрасли“ добро репрезентује ове групе.

Како бисмо укључили пикове младунаца у категорију младих, категорија младих ће обухватати све абалоне са 8 и мање прстена, а чланове одраслих и (делове младунаца) ћемо поделити на зреле и старе. У поставци задатка наведено је да је број прстена довољан податак да бисмо одредили старост, па категорије можемо поделити на једнаке делове без обзира на укупну расподелу броја прстена. Како смо претходно утврдили да најзначајнији број абалонеа има до 23 прстена и следећа категорија „зрели“ ће обухватити 8 прстена. Категорија „стари“ ће обухватити осталих 7 прстена као и остатак који ће на неки начин представљати допуну до осмог прстена. Старосна категорија биће додата у новој колони (слика 17), а колона са бројем прстена биће избачена из *data frame*-а (слика 18).

   
***Слика 17.*** *Подела на старосне категорије*

   
***Слика 18.*** *Избацивање прстена*

У кораку 3.1. већ смо обавили одређену обраду података када смо избацили очигледне грешке и кодирали називе полова.

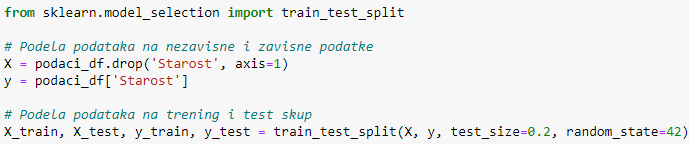
# 4. Неуронска мрежа

## 4. 1. Подела скупа података

Пре него што направимо неуронску мрежу потребно је поделити скуп података. Дефинисаћемо нове скупове и то:

* Тренинг скуп – који ће садржати 80% података
* Тест скуп – који ће садржати 20% података

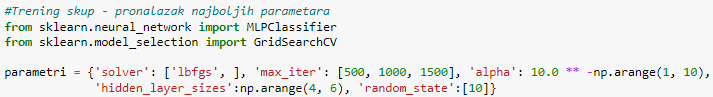
За чување оба скупа биће потребне по две променљиве. Прва променљива ***X*** чуваће улазне, односно независне податке, док ће друга променљива ***y*** чувати излазне, односно зависне променљиве (слика 19).

   
***Слика 19.*** *Подела података на тренинг и тест скуп*

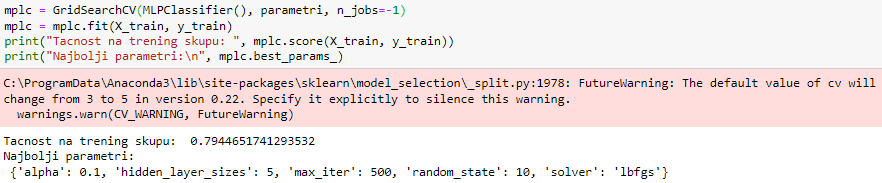
## 4. 2. Избор најбољих параметара

За решавање задатка коришћен вишеслојни перцептрон класификатор (енг. *Multi-layer Perceptron classifier*) - ***MLPClassifier***.

Овај класификатор примењује алгоритам вишеслојног перцептрона који тренира мрежу користећи пропагацију грешке уназад (енг. *Backpropagation*). Предложени су следећи параметри (слика 20):

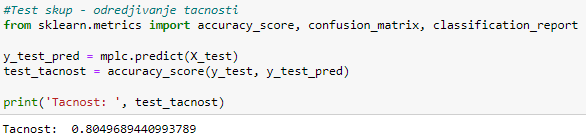
  
***Слика 20.*** *Тренирање мреже различитим комбинацијама параметара*

Функција ***GridSearchCV*** пролази кроз све могуће комбинације задатих параметара, а када се позове функција ***fit*** (слика 21) враћа најбољу могућу комбинацију.

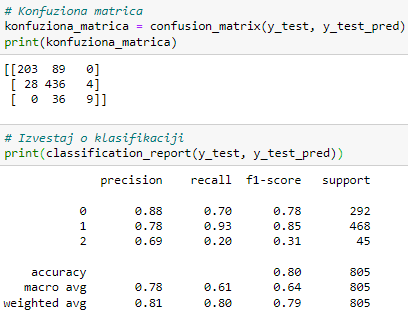
  
***Слика 21.*** *Враћање најбоље комбинације и штампање тачности на тренинг скупу*

## 4. 3. Тестирање мреже

Тачност мреже на тест скупу дата је испод (слика 22).

  
***Слика 22.*** *Тачност на тест скупу*

На слици испод приказана је конфузиона матрица и извештај о класификацији (слика 23).

***Слика 23.*** *Конфузиона матрица и извештај о класификацји*

# 5. Закључак

На основу добијених резултата прецизности на тест скупу, конфузионе матрице и извештаја о класификацији може се рећи да је неуронска мрежа успешна у класификацији задатог проблема.

Најмањи проценат погодака има на класи број 2, односно на старим абалонима. Овој класи између осталих припадају јединке које су у поставци имале више од 23 прстена. Како смо раније видели за сваки прстен већи од 23 постојала је само једна или две јединке, па је вероватно постојала грешка при прикупљању ових података.

# 6. Литература

1. Курс: Вештачка интелигенција, доступно на: <http://moodle.fink.rs/course/view.php?id=989>, приступљено 28.05.2023.]
2. Сајт: Anaconda Documentation, доступно на: [https://docs.anaconda.com](https://docs.anaconda.com/), приступљено 28.5.2023.]
3. Сајт: Jupyter Project Documentation, доступно на: <https://docs.jupyter.org>, приступљено 28.5.2023.]
4. Сајт: 3.11.3 Documentation, доступно на: [https://docs.python.com](https://docs.python.com/), приступљено 28.5.2023.]
5. MLPClassifier, доступно на: <https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html> , приступљено 28.05.2023.]