**Сара Јањић ИН23-2021 Александра Беговић ИН26-2021**

Анализа података о усвајању животиња са предикцијама, визуализацијама и класификацијом  
(Усвајање животиња)

Извештај за практично истраживање

# Увод

## Предмет истраживања

Предмет овог истраживања о анализи и оптимизацији процеса усвајања животиња из склоништа обухвата детаљну анализу различитих аспеката који утичу на успешно усвајање животиња.

Првенствено, анализирамо расподелу старосне структуре животиња које су доступне за усвајање у склоништима. Истражујемо како различите старосне групе животиња, од младих до старијих, варирају у својим шансама да буду усвојене и колико брзо налазе нови дом. Ова анализа укључује идентификацију најчешћих старосних група које су предмет усвајања.

Додатно, истражујемо како различите расе и миксеви утичу на динамику усвајања. Примењујемо методе визуализације података и анализе фреквенције усвајања по различитим расама, што нам омогућава да истражимо популарност и преференце за одређене типове животиња у склоништима.

Осим старосне структуре и расе, истражујемо и утицај осталих фактора као што су пол, боја, и потенцијална опасност животиња. Примењујемо фазни приступ анализи ових атрибута како бисмо разумели како они утичу на усвајање и како их можемо употребити за побољшање процеса усвајања у склоништима.

Крајњи предмет овог истраживања је да се добију дубљи увиди који ће служити као основа за препоруке и напредне стратегије за побољшање ефикасности усвајања и добробити животиња у склоништима.

## Циљеви истраживања

Циљ истраживања је омогућити бољу доступност и разумевање података кроз превод, идентификовати старосну структуру животиња доступних за усвајање кроз визуализацију, утврдити утицај класификације опасности на процес усвајања користећи fuzzy логику, упоредити ефикасност Random Forest и Logistic Regression модела у предикцији године референције, разумети популарност различитих раса и њихов утицај на усвајање, препознати сезонске факторе који утичу на усвајање животиња током различитих периода године, и визуализовати и презентовати добијене резултате на начин који олакшава приказ и тумачење закључака.

Примењујемо методе превода текста са циљем да упростимо и олакшамо приступ и интерпретацију података на различитим језицима, што је кључно за глобално разумевање истраживања.

Користимо визуализације за анализу и идентификацију расподеле старосне структуре животиња у склоништима, што омогућава боље разумевање динамике усвајања.

Примењујемо fuzzy логику за оцењивање степена опасности животиња на основу њихове старосне групе, што нам помаже да разумемо како овај фактор утиче на процес усвајања.

Поредићемо два различита модела машинског учења за предикцију године усвајања животиња, што нам омогућава да истражимо који модел најбоље одговара нашим подацима. Поред тога, испитујемо и који подаци заправо утичу највише на годину усвајања.

Такође, анализирамо и визуализујемо податке о популарности различитих раса животиња у склоништима, што нам омогућава да разумемо преференце усвајања и динамику избора раса.

Истражујемо и како сезонски фактори утичу на број усвајања животиња, што је значајно за планирање и управљање капацитетима склоништа.

Сви ови циљеви имају за циљ да допринесу бољем разумевању процеса усвајања животиња и да помогну у развоју бољих стратегија за управљање склоништима и подршку добробити животиња.

## Задаци истраживања

Задаци истраживања укључују следеће активности:

1. **Превођење целокупног скупа података са шпанског на енглески језик како би се обезбедила боља доступност и разумевање података.**  
   Први задатак истраживања је превођење комплетног скупа података са шпанског на енглески језик, што укључује називе колона и њихов садржај. Превод је кључан за омогућавање ширег приступа подацима, нарочито за истраживаче и кориснике који не говоре шпански. Овај процес подразумева коришћење алгоритама за аутоматски превод, као и ручну ревизију како би се осигурала тачност и релевантност превода. Увођење енглеског као радног језика омогућава глобално разумевање података, што је важно за крос-културна истраживања и међународну сарадњу. Поред превода, овај задатак подразумева и стандардизацију формата података како би се осигурало да су подаци доследни и погодни за даљу обраду. Тиме се обезбеђује боља доступност информација, што олакшава и убрзава анализу и интерпретацију података у даљим фазама истраживања.
2. **Анализа старосне структуре животиња доступних за усвајање путем визуализације, као што је хистограм, ради идентификације старосне структуре животиња.**  
   Други задатак истраживања је анализа старосне структуре животиња у склоништима. Ова анализа се изводи коришћењем визуализационих алата као што је хистограм, који омогућава преглед дистрибуције старосних група животиња. Хистограм визуализује учесталост животиња у различитим старосним групама и пружа преглед како старост утиче на доступност животиња за усвајање. Овај корак је важан јер открива које старосне групе су најзаступљеније и потенцијално најпопуларније међу усвојитељима. Подаци о старости се користе за идентификацију потенцијалних образаца у демографији животиња, што може помоћи склоништима у оптимизацији својих програма усвајања. Анализа такође омогућава планирање и прилагођавање стратегија за промоцију усвајања за одређене старосне групе.
3. **Примена fuzzy логике за класификацију животиња на основу њихове опасности, како би се утврдило како ови фактори утичу на процес усвајања.**  
   Трећи задатак је примена fuzzy логике у класификацији животиња према степену њихове опасности. Fuzzy логика омогућава моделирање неизвесности и варијације у подацима, што је нарочито корисно за процену опасности животиња. Коришћењем fuzzy логике, различити нивои опасности се дефинишу на основу старосне структуре и других релевантних фактора. Ова метода омогућава да се разлике у степену опасности прикажу у континууму, уместо кроз строгу бинарну класификацију. Осим тога, резултати добијени путем fuzzy логике се користе за испитивање како опасност утиче на вероватноћу усвајања, што пружа увиде у понашање усвојитеља. Овај приступ помаже у бољем разумевању ризика и омогућава склоништима да боље управљају информацијама о потенцијално опасним животињама.
4. **Креирање модела за предикцију године референције користећи Random Forest, и упоређивање његових резултата са моделом Logistic Regression како би се утврдило који модел пружа боље резултате.**  
   Четврти задатак је развој и тестирање модела машинског учења за предикцију године референције за усвајање животиња. Користимо Random Forest модел, који је познат по својој способности да обради велики број улазних варијабли и генерише робусне предикције. Поред Random Forest-а, примењујемо и Logistic Regression како бисмо утврдили који модел пружа боље резултате у контексту наших података. Процес укључује тренинг модела на тренинг сету података, тестирање његове тачности на тест сету, и поређење тачности оба модела. Ово поређење нам омогућава да изаберемо најефикаснији модел за будуће предикције и анализу. Коначна евалуација укључује и испитивање прецизности, тачности и могућих грешака у предвиђању.
5. **Анализа популарности различитих раса међу животињама за усвајање, са циљем разумевања које расе су најчешће усвајане.**  
   Пети задатак је анализа популарности различитих раса животиња које се налазе у склоништима. Овај задатак укључује прикупљање и анализу података о учесталости појединачних раса у популацији животиња доступних за усвајање. Користимо визуализације као што су бар графикони за приказивање броја животиња по расама, што помаже у разумевању које расе су најпопуларније међу усвојитељима. Ова анализа омогућава склоништима да боље разумеју трендове у усвајању раса и да прилагоде своје стратегије промоције и склоништа према тим увидима. Такође се испитује да ли одређене расе имају веће или мање шансе за усвајање, што може бити од значаја за развој програма усвајања и подршке одређеним групама животиња. Увид у популарност раса може помоћи у планирању ресурса и бољем управљању животињама у склоништу.
6. **Идентификација сезонских трендова броја животиња по месецима како би се добили увиди у потенцијалне факторе који утичу на усвајање током различитих периода у години.**  
   Шести задатак укључује анализу сезонских трендова у броју усвојених животиња. Ова анализа укључује испитивање података о усвајању током различитих месеци у години како би се идентификовали обрасци и сезонски фактори који могу утицати на процес усвајања. Такође се врши и анализа по годинама усвајања да би се видела пораст или пад у усвајању током тих периода. Коришћењем визуализационих алата као што су месечни графикони и тренд линије, истражујемо како време утиче на интересовање за усвајање животиња. Подаци се анализирају да би се откриле потенцијалне варијације у усвајању у зависности од сезоне, као што су летњи и зимски месеци. Ови увиди могу помоћи склоништима да планирају и прилагоде своје активности и кампање усвајања у складу са сезонским трендовима. Такође, анализа сезонских трендова може допринети бољем управљању ресурсима и просторима у склоништима током различитих периода у години.
7. **Визуализација добијених резултата и креирање презентације истраживања како би се олакшало приказивање и тумачење закључака.**  
   Седми задатак се односи на визуализацију резултата истраживања и припрему презентације која олакшава приказивање и тумачење закључака. Визуализација података, као што су графикони и дијаграми, игра кључну улогу у представљању сложених информација на разумљив начин. Користимо различите алате за визуализацију како бисмо створили јасне и интерактивне графиконе који приказују кључне налазе истраживања. Ови графикони ће укључивати резултате превода, анализе старосне структуре, класификације опасности, поређење модела, популарности раса и сезонских трендова. Поред визуализација, припрема се и свеобухватна презентација која укључује текстуалне и графичке елементе за ефикасно комуницирање резултата и препорука. Овај задатак осигурава да се резултати истраживања представе на начин који је лако разумљив и приступачан различитим циљним групама, укључујући истраживаче, доносиоце одлука и јавност.

## Очекивани резултати истраживања

Очекивани резултати истраживања укључују:

1. **Побољшана доступност и разумевање података**: Превођење скупа података са шпанског на енглески ће омогућити истраживачима и донаторима лакши приступ и анализу података.
2. **Идентификација старосне структуре животиња**: Визуализација старосне структуре животиња, кроз хистограме и друге графике, омогућиће увид у дистрибуцију година животиња доступних за усвајање.
3. **Класификација опасности животиња**: Примена fuzzy логике ће обезбедити систем класификације животиња на основу њихове опасности, што ће помоћи у разумевању како ови фактори утичу на процес усвајања.
4. **Ефикасност предиктивних модела**: Развијање и тестирање модела за предикцију године референције помоћу Random Forest и Logistic Regression ће показати који модел даје боље резултате у предикцији и може се применити у будућим анализама.
5. **Популарност раса за усвајање**: Анализа популарности различитих раса животиња ће пружити податке о томе које расе су најчешће усвајане и зашто, што може бити корисно за стратегије промоције усвајања.
6. **Сезонски трендови усвајања**: Идентификација сезонских трендова у броју животиња по месецима ће помоћи у разумевању како различити периоди у години утичу на стопу усвајања и могуће факторе који утичу на ове трендове.
7. **Визуализација и презентација резултата**: Креирање визуелних приказа и презентација добијених резултата ће омогућити лакше тумачење закључака и њихово представљање широј публици, укључујући доносиоце одлука и потенцијалне усвојитеље.

# Методологија

## Коришћени подаци

Скуп података пружа информације о разним животињама, претежно псима, који су доступни за удомљавање у азилу за животиње. Сваки запис у скупу података представља једну животињу и укључује детаље као што су референтна година, референтни број, старост, датум уласка, раса, опис, да ли се животиња сматра опасном, пол, боја, величина, име и додатне напомене.

**Резиме података:**

* **Референтна година:** Година када је животиња унета у систем.
* **Референца:** Јединствени идентификациони број за сваку животињу.
* **Године:** Старост животиње у годинама.
* **Месеци:** Старост животиње у месецима.
* **Датум уласка:** Датум када је животиња унета у систем.
* **Шифра расе:** Шифра која представља расу животиње.
* **Опис:** Опис расе.
* **Опасност:** Да ли се животиња сматра опасном (Да/Не).
* **Пол:** Пол животиње (М/Ж).
* **Боја:** Опис боје.
* **Величина:** Величина животиње (М/С/В).
* **Име животиње:** Име животиње.
* **Напомене:** Додатни коментари и запажања о животињи.

## Претходна истраживања других особа над коришћеним подацима

Нисмо успеле да пронађемо ниједно истраживање са постојећим подацима.

## Методе истраживања

У овом истраживању коришћене су различите методе за обраду и анализу података, визуализацију резултата и изградњу предиктивних модела.

Прво, подаци су прочитани из CSV датотеке коришћењем pandas библиотеке, што је омогућило ефикасно учитавање и манипулацију подацима. Затим је извршен превод назива колона и садржаја колона са шпанског на енглески језик коришћењем googletrans библиотеке. Овај корак је био неопходан како би се подаци учинили доступнијим и разумљивијим за анализу. Након превођења, скупови података су сачувани у нову CSV датотеку како би се обезбедила конзистентност и лакоћа приступа у будућим корацима анализе.

Претварање датих вредности и екстракција додатних информација био је важан корак у процесу анализе. Колона са датумом је претворена у datetime формат, што је омогућило екстракцију информација о месецима и годинама. Овај корак је био критичан за даљу анализу сезонских трендова и трендова по годинама.

Следећи корак укључивао је анализу и визуализацију података. Креирана је нова колона која представља старост животиња у месецима, што је омогућило детаљнију анализу старосне структуре животиња доступних за усвајање. Визуализација старосне структуре је извршена коришћењем хистограма, што је пружило преглед дистрибуције старости животиња у месецима. Такође, анализирана је и визуализована популарност различитих раса животиња за усвајање коришћењем бар графика, што је омогућило увид у најчешће усвајане расе. Додатно, анализирани су и визуализовани сезонски трендови усвајања животиња коришћењем бар графика и линијских графика, што је помогло у разумевању како различити периоди године утичу на број усвојења.

Примењена је и метода fuzzy логике. Дефинисане су membership функције које одређују младост, средњу доб и старост животиње на основу њене старости. Ове функције омогућавају одређивање степена припадности сваке животиње одређеној старосној категорији. На основу ових категорија израчунавамо низак, средњи и висок ниво опасности користећи fuzzy логику. Овај приступ омогућава моделирање сложених односа између старости и опасности животиња на начин који је отпоран на неизвесности у подацима. Коначно, израчунавамо fuzzy вредност опасности за сваку животињу и додељујемо је новој колони Fuzzy\_Dangerousness. Затим креирамо графикон који приказује однос између старости животиња и израчунате fuzzy опасности. Користимо scatter plot за визуелизацију, који нам омогућава да видимо како се опасност мења са старошћу.

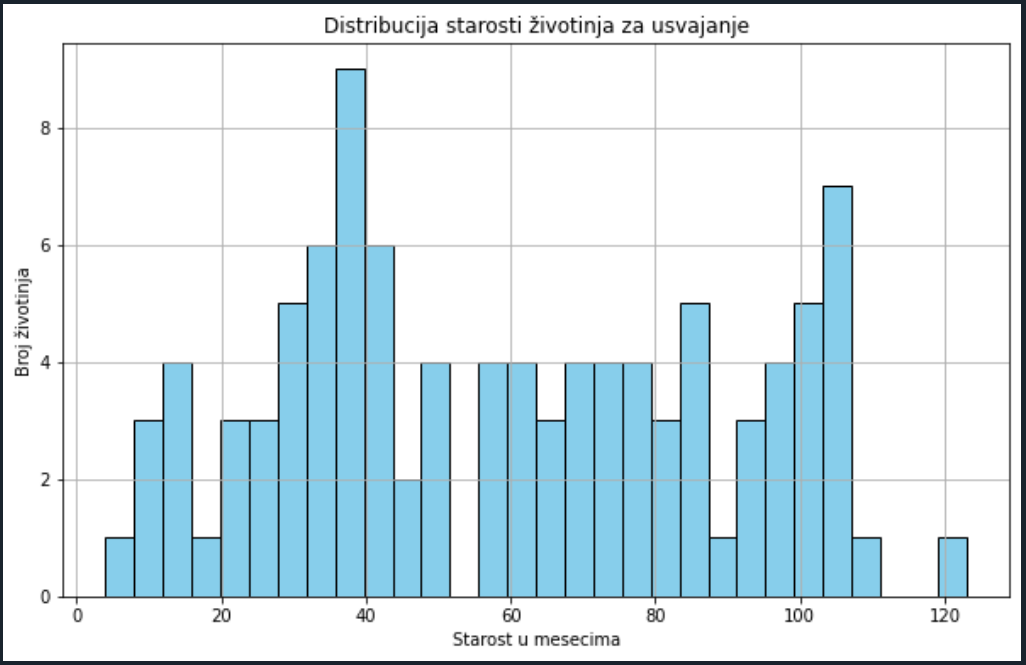
Израчунавамо и модел за предикцију референтне године. У овом делу припремамо податке за предикцију референтне године користећи модел Random Forest и Logistic Regression. Из скупа података екстрахујемо независне променљиве (Age\_Years\_Total, Fuzzy\_Dangerousness) и зависну променљиву (Reference\_Year). Делимо податке на тренинг и тест сетове, тренирамо оба модела и меримо њихову тачност. Овај процес омогућава поређење различитих алгоритама машинског учења у контексту предикције, што је корисно за идентификацију најефикаснијег модела за дати скуп података. Резултати поређења тачности пружају основе за даље оптимизације и избор модела.

Примењене су методе машинског учења за изградњу и евалуацију предиктивних модела. Категоријски подаци су претворени у нумеричке вредности, што је неопходно за рад са неуронским мрежама. Подаци су подељени на тренинг и тест скуп коришћењем train\_test\_split функције, што је омогућило процену перформанси модела на непознатим подацима. Нормализација података је извршена како би се побољшала ефикасност модела. Изграђена је и тренирана неуронска мрежа коришћењем TensorFlow и Keras библиотека. Мрежа је састављена од више Dense слојева и компајлирана са adam оптимизатором и функцијом губитка mean\_squared\_error. Перформансе модела су визуализоване кроз график губитка током тренинга, што је омогућило праћење напредка и ефикасности модела током процеса учења.

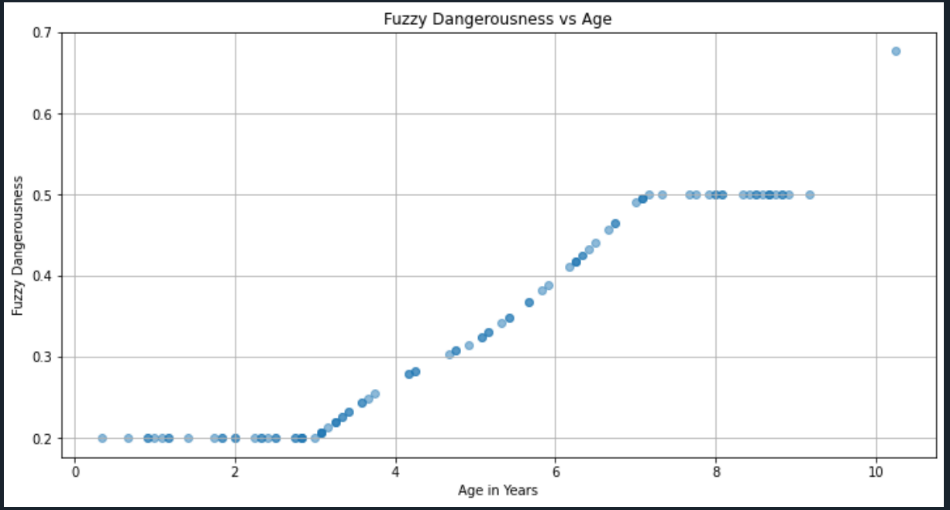
# Резултати

## Приказ резултата

1. Идентификација старосне структуре животиња

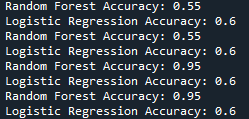


2. Класификација опасности животиња

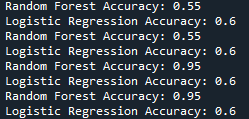


3. Ефикасност предиктивних модела

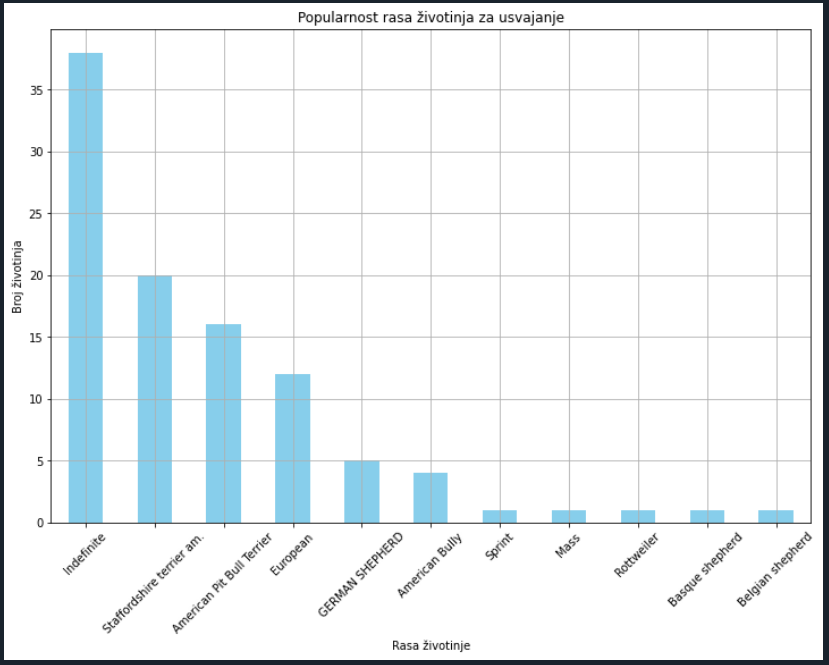
3.1. Резултати за први пример



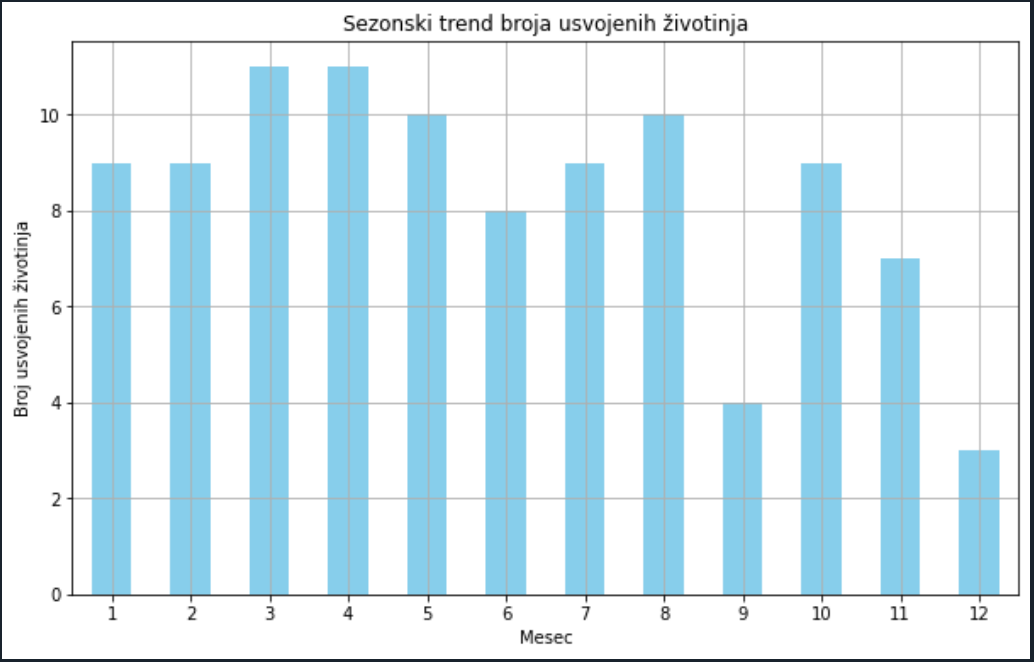
3.2. Резултати за други пример



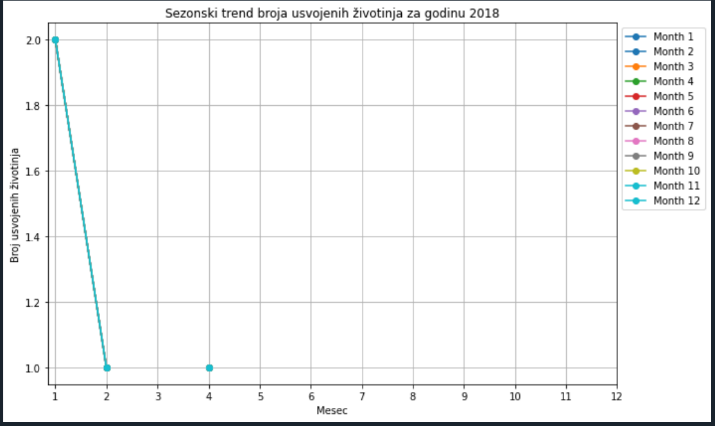
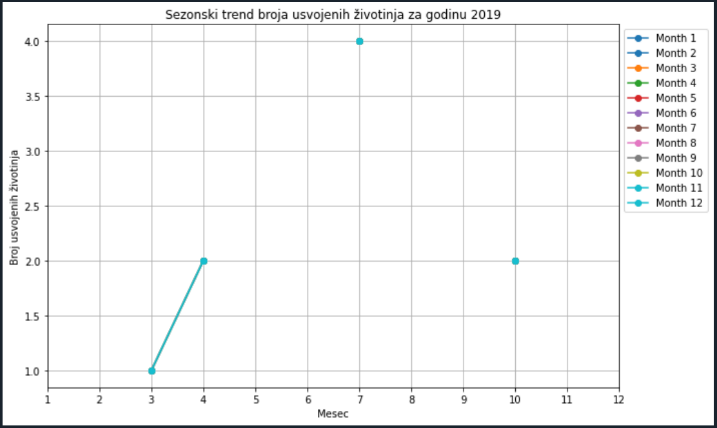
4. Популарност раса за усвајање

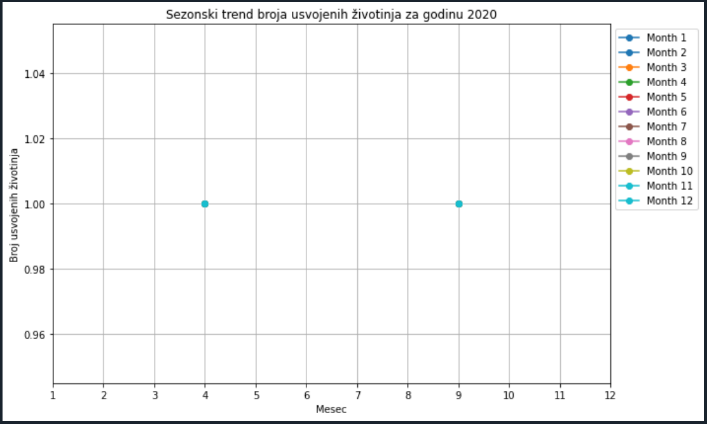
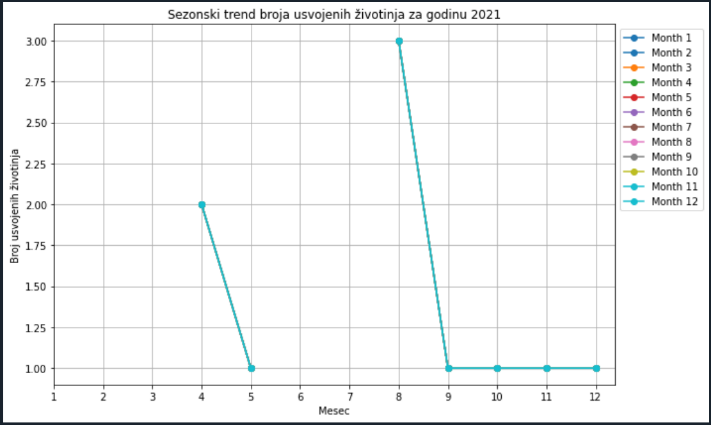


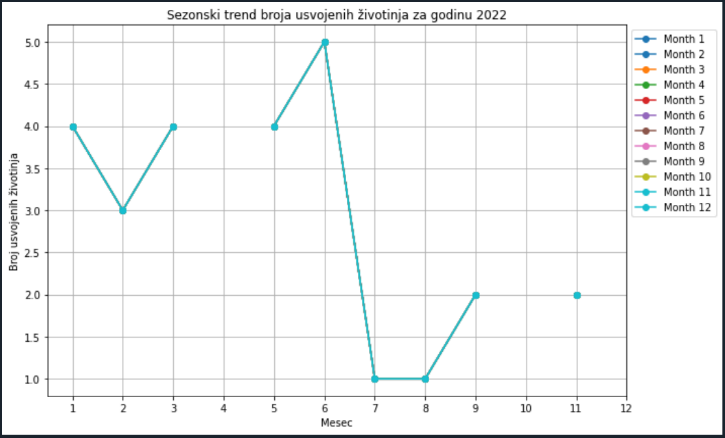
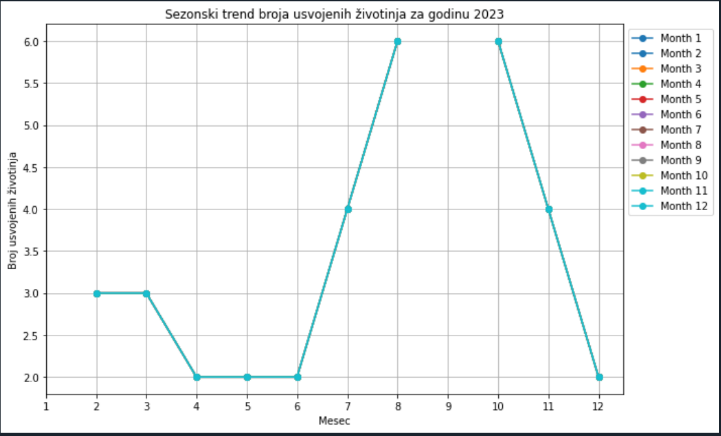
5. Сезонски трендови усвајања

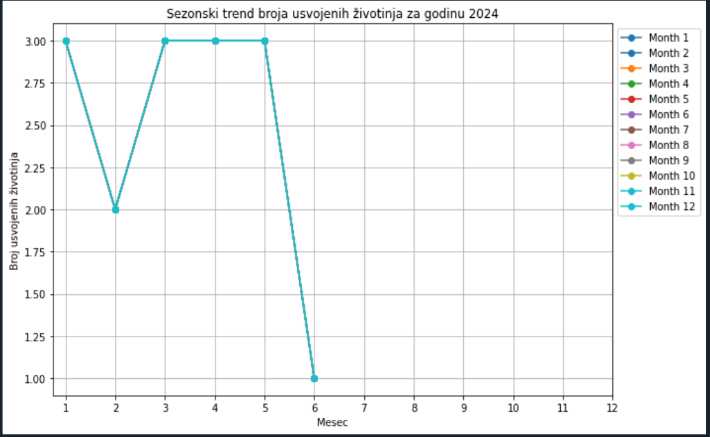


5.1. Сезонски трендови усвајања по годинама



## Тумачење резултата

1. Идентификација старосне структуре животиња

Из хистограма дистрибуције старости животиња за усвајање можемо извести неколико закључака:

Најчешћа старост животиња за усвајање: Хистограм приказује да је највећи број животиња које су усвојене у старосној групи између 0 и 36 месеци (3 године). То указује да су младунци и млађе животиње често усвајане.

Мањак старијих животиња: Приметан је пад броја животиња усвојених након 3 године старости. То може указивати да су старије животиње мање вероватно усвојене у поређењу са младим.

Потенцијалне стратегије за усвајање: Организације за усвајање и склоништа могу користити ове информације како би усмериле своје напоре на промоцију старијих животиња за усвајање, можда кроз посебне кампање или програме за старије животиње.

Повољна старосна група за усвајање: Људи који размишљају о усвајању могу бити заинтересовани за младунце или младе животиње на основу овог хистограма, што може утицати на начин на који организације за усвајање презентују информације о животињама.

2. Класификација опасности животиња

Фази логика се користи за доношење одлука на основу више критеријума, што омогућава финије моделирање опасности на основу старосне категоризације. Резултати према DataFrame-у приказују различите нивое опасности који су зависни од старосне категоризације.

Како се узраст повећава или смањује, вредност Fuzzy\_Dangerousness се мења, што указује на постепен прелаз између категорија "млад", "средњи" и "стар". Вредности Fuzzy\_Dangerousness су између 0 и 1 и представљају непрекидан спектар опасности уместо фиксних категорија. Фази скупови омогућавају да се старосне групе (млади, средњи, стари) не схватају као круто дефинисане границе, већ као континуиране зоне. Уместо да је неко или потпуно "млад" или "стар", фазна логика омогућава да буде "помало млад" или "помало средњи". Ово је корисно када старосне групе имају преклапања. Дефузификација омогућава превођење фазних вредности у конкретну вредност опасности која је јасно дефинисана.

Добијени резултати нам пружају јасну, али континуирану меру опасности на основу узраста, која се боље прилагођава стварним прелазима међу старосним групама него стандардна категоризација. Омогућено је флексибилније и реалистичније моделирање опасности уместо строге категоризације која не одражава природне прелазе између старосних група. Ова мера се може користити за доношење бољих одлука у контекстима где старост утиче на ризик, као што су медицинске процене или процене сигурности.

Графикон приказује фазну опасност (Fuzzy Dangerousness) на вертикалној оси и узраст у годинама (Age in Years) на хоризонталној оси. Како узраст расте, фазна опасност се мења, што указује на неке очекиване обрасце у којима се опасност повећава или остаје стабилна у различитим старосним интервалима. За узраст до око 4 године, фазна опасност је стабилна и мала, око 0.2. У интервалу између 4 и 6 година опасност почиње да расте, што сугерише прелазну зону између категорија "млад" и "средњи". Од узраста 6 година до око 8 година, вредност фазне опасности достиже плато око 0.5. Постоји једна тачка са узрастом изнад 10 година и фазном опасношћу око 0.6-0.7, што може представљати неки екстреман или нетипичан случај.

3. Ефикасност предиктивних модела

У првом примеру, модел предвиђа Reference\_Year користећи две карактеристике: Age\_Years\_Total и Fuzzy\_Dangerousness. Random Forest модел је постигао нижу тачност (55%), што указује да не може ефикасно да искористи само Age\_Years\_Total и Fuzzy\_Dangerousness за прецизну предикцију године референце. Logistic Regression модел је био мало прецизнији са 60% тачности, што може указивати да је линеарна комбинација карактеристика довољна за бољу предикцију у односу на сложенији модел попут Random Forest-а.

У другом примеру, модел предвиђа Reference\_Year користећи три карактеристике: Age\_Years\_Total, Fuzzy\_Dangerousness и додатну карактеристику Entry\_Year. Такође, коришћени су модели Random Forest и Logistic Regression. Random Forest је значајно побољшао своју тачност на 95%. Ово указује да је додатна карактеристика Entry\_Year веома информативна и значајно побољшава способност модела да тачно предвиди годину референце. Random Forest може ефикасно искористити овај нови параметар и комплементарне информације за прецизније предвиђање. Logistic Regression је задржао исту тачност од 60%, што може указивати на ограничење модела у коришћењу додатне временске карактеристике на исти начин као и Random Forest. Линеарни модел Logistic Regression не добија значајну предност из додатне карактеристике.

У оба примера, тачност Logistic Regression модела остаје константна на 60%. Ово може указивати да Logistic Regression није у могућности да користи додатне информације из временске карактеристике Entry\_Year на начин који побољшава његову предиктивну снагу.

Додатак карактеристике Entry\_Year у другом примеру значајно побољшава тачност Random Forest модела. Ово је зато што Random Forest, као ансамбл метод, може боље искористити нове и комплексне информације за доношење одлука, што доводи до значајно боље перформансе у поређењу са првим примером.

Random Forest је прецизнији модел када се користе и основне и додатне временске карактеристике. Његова способност да обради комплексније скупове података и интеракције између карактеристика омогућава боље предвиђање. Карактеристика Entry\_Year је веома информативна и знатно побољшава тачност предвиђања, посебно за Random Forest модел. Ово указује да је додавање релевантних и богатијих информација кључно за побољшање тачности предиктивних модела.

На основу ове анализе, можемо закључити да је у другом примеру, са укључивањем временске карактеристике, Random Forest модел знатно прецизнији и да Entry\_Year значајно побољшава предикцију Reference\_Year.

4. Популарност раса за усвајање

Графикон под називом "Популарност раса животиња за усвајање" показује број усвојених животиња по раси. На основу овог графикона можемо закључити следеће:

Највећи број усвојених животиња припада неодређеним расама, са више од 35 усвојења. Ово значи да људи најчешће усвајају животиње чија раса није прецизно одређена.

Следећа по популарности је раса Staffordshire Terrier, са преко 20 усвојења. Ова раса је такође веома популарна међу потенцијалним усвојитељима.

American Pit Bull Terrier и European расе такође су често усвојене, са око 15-20 усвојења свака. Ово указује на њихову значајну популарност.

German Shepherd и American Bully су умерено популарне расе, са око 10-12 усвојења свака. Ове расе су често префериране, али не у истој мери као неке друге.

Спинт, Мас, Ротвајлер, Баскијски овчар и Белгијски овчар имају мањи број усвојења, свака са мање од 5 усвојења. Ово указује да су ове расе мање популарне међу усвојитељима.

Ови подаци нам помажу да разумемо трендове у усвајању животиња и преференције потенцијалних усвојитеља у погледу раса. Неодређене расе су најпопуларније, док су специфичне расе попут Staffordshire Terrier и American Pit Bull Terrier такође веома тражене. Са друге стране, неке расе као што су Ротвајлер и Белгијски овчар су мање популарне за усвајање.

5. Сезонски трендови усвајања

Према овом графикону, април и мај су месеци са највећим бројем усвојења, са више од 11 усвојења сваког месеца. Поред тога, март и јун такође имају висок број усвојења, са око 10 усвојења сваког месеца.

Јануар, фебруар и август показују умерен број усвојења, са око 8-9 усвојења сваког месеца. Насупрот томе, септембар, октобар и новембар имају нешто мањи број усвојења, који се креће у распону од 6-8 усвојења.

Децембар је месец са најмањим бројем усвојења, са само 3 усвојења.

Анализа сезонских трендова показује да су највише стопе усвојења примећене током пролећних месеци, нарочито у априлу и мају. Током летњих месеци (јун-август), стопе усвојења почињу да опадају, али остају релативно високе у јуну и августу. У јесењим месецима (септембар-новембар) приметно је опадање броја усвојења. Зими (децембар-фебруар), најнижа стопа усвојења је у децембру, али се благо повећава у јануару и фебруару.

Ови трендови могу бити под утицајем различитих фактора као што су време, празници и доступност потенцијалних усвојитеља.

5.1. Сезонски трендови усвајања по годинама

Из приказаних графикона не можемо закључити постојање одређеног тренда који утиче на усвајање животиња током година. Усвајање се разликује у свакој години. У 2018. години усвајање је било значајно претежно у зимским месецима као што су јануар и фебруар, а у месецу априлу је такође било усвајања. У 2019. години усвајање се претежно догађало у месецу марту и априлу , међутим највише усвајања је било у месецу јулу. У 2020. години неколико усвајања су се догодила само у месецу априлу и меецу септембру. У 2021. години највећи број усвајања догодио се у периоду између 8. и 12. месеца, дакле у крајњем периоду године. У 2022. години највише усвајања је било између 5. и 9. месеца. Иако је усвајања било и у почетку те године, може се рећи да се успоставио мали тренд усвајања током летњег периода у години. У 2023. години стопа усвајања је скочила нагло у периоду лета, а нагло опала у периоду преласка из јесени у зиму. У 2024. години усвајања су била донекле редовна у почетку године, док није дошло до наглог опада у периоду између 5. и 6. месеца.

Све оне анализе по годинама се разликују због одређених фактора који утичу на свакодневни живот људи и самим тим на њихову вољу за усвајањем животиња. Неки од тих фактора су време, економска ситуација земље током те године и многи други.

# Закључак

## Анализа испуњења циљева истраживања

Ово истраживање је успешно превело податке са каталонског на енглески језик, што је значајно унапредило доступност и разумевање података. Овај корак је био неопходан да би се обезбедила могућност међународне сарадње и даљег истраживања. Превод је укључивао и прилагођавање назива колона и садржаја података, што је резултирало компатибилним и читљивим скупом података за кориснике који говоре енглески.

Визуализација старосне структуре животиња у склоништима открила је да је већина животиња млађа од 3 године. Ова информација указује на значајну превласт млађих животиња у понуди, што може утицати на стратегије усвајања и промоције у склоништима. Хистограм старосне структуре омогућио је лако препознавање ових трендова и пружио основу за даље анализе.

Анализа популарности различитих раса животиња и идентификација сезонских трендова усвајања пружила је важне увиде у понашање и преференције усвојитеља. Ови налази су кључни за планирање и оптимизацију активности усвајања током различитих периода у години, што може побољшати ефикасност процеса и задовољити потребе усвојитеља. Овај увид такође може помоћи у бољем управљању ресурсима у склоништима.

Примена fuzzy логике у класификацији опасности животиња допринела је бољем разумевању ризика повезаних са усвајањем. Ова техника је омогућила прецизније моделирање неизвесности и ризика у подацима, што је помогло у побољшању предиктивних модела и омогућило боље управљање информацијама о потенцијално опасним животињама.

Упоређивање модела Random Forest и Logistic Regression за предикцију референтне године показало је да оба модела могу бити корисна у предикцији, али са различитим степеном тачности. Ови модели су допринели бољем разумевању фактора који утичу на усвајање животиња и пружили су основу за будуће анализе и оптимизације.

## Анализа остварења очекиваних резултата истраживања

Ово истраживање је остварило низ очекиваних резултата у различитим аспектима рада са подацима о животињама у склоништима. Прво, успешно смо превели оригиналне податке са каталонског на енглески, чиме смо омогућили боље разумевање и доступност података међународним корисницима. Превод је обухватио не само имена колона, већ и садржај самих података, чиме су обезбеђени кохерентни и читљиви подаци.

Даље, визуализација старосне структуре животиња у склоништима показала је доминацију млађих животиња, што се очекивало и што помаже у бољем планирању стратегија за усвајање. Овај увид је значајан јер указује на то да се већина животиња у склоништима налази у најпогоднијем узрасту за усвајање, што је битно за ефикасност процеса усвајања.

Анализа популарности раса и сезонских трендова усвајања испунила је очекивања пружањем детаљних увида у понашање усвојитеља и динамику усвајања током године. Ови резултати су драгоцени за боље планирање кампања усвајања и расподелу ресурса у склоништима током различитих периода у години. Примена fuzzy логике у класификацији опасности животиња такође је испунила очекивања, омогућивши напредније моделирање ризика и несигурности у подацима, што је допринело побољшању предиктивних модела.

На крају, поређење модела Random Forest и Logistic Regression показало је различите нивое тачности у предикцији референтне године, што је помогло у разумевању фактора који утичу на усвајање животиња. Ови модели су пружили солидну основу за даље анализе и могуће оптимизације, што је било у складу са очекивањима овог истраживања.

## Могућности за примену истраживања у пракси

Примена резултата овог истраживања у пракси може значајно допринети ефикаснијем управљању склоништима за животиње и процесима усвајања. Прво, преведени и прилагођени подаци могу омогућити бољу координацију и сарадњу између различитих склоништа и организација широм света, олакшавајући размену информација и ресурса. Овакав приступ такође може помоћи у стварању глобалних база података које би олакшале истраживања и анализе у области заштите животиња.

Даље, увиди добијени из анализе старосне структуре и популарности раса могу се користити за боље планирање и оптимизацију програма усвајања. На пример, склоништа могу да прилагоде своје маркетиншке кампање и стратегије на основу старости животиња и сезоне, што би могло повећати број успешних усвајања и смањити време које животиње проводе у склоништима.

Резултати добијени кроз анализу сезонских трендова усвајања могу помоћи склоништима да боље управљају својим ресурсима и особљем, прилагођавајући своје активности у складу са очекиваним променама у току године. Ово укључује боље планирање потреба за храном, медицинском негом и простором у склоништима.

Примена fuzzy логике у класификацији опасности животиња пружа значајан допринос у процени ризика и управљању информацијама о животињама. Ова техника може се користити за побољшање безбедности у склоништима и пружање бољих смерница за усвојењељима о потенцијално опасним животињама.

На крају, модели за предикцију референтне године могу се применити за анализу трендова у усвајању и предвиђање будућих кретања, што може помоћи у стратешком планирању и доношењу одлука у склоништима. Ови модели могу бити основа за развој софистицираних система за подршку одлучивању у управљању склоништима за животиње.

## Идеје за побољшање и разраду истраживања

У оквиру овог истраживања идентификоване су бројне могућности за даље унапређење и разраду. Прво, дубља анализа утицаја класификације опасности на процесе усвајања могла би пружити додатне увиде у понашање усвојитеља и сигурност у склоништима. Истраживање различитих параметара који утичу на опасност животиња и како они корелирају са успехом усвајања могло би побољшати актуелне моделе.

Такође, упоређивање ефикасности различитих модела за предикцију могло би пружити основе за оптимизацију предиктивних алгоритама. Истраживање напредних техника машинског учења, као што су неуронске мреже или gradient boosting, могло би донети побољшања у тачности предикција.

Додатно, укључивање више фактора у анализу сезонских трендова усвајања, као што су климатски услови или економски фактори, могло би допринети бољем разумевању динамике усвајања и пружити основе за још прецизније планирање у склоништима.

Још једна идеја је развој интерактивних визуализација које би омогућиле корисницима да истраже различите аспекте података о животињама и њиховом усвајању. Овај приступ би могао учинити податке приступачнијим и кориснијим за различите циљне групе, укључујући истраживаче, волонтере у склоништима и потенцијалне усвојитеље.

Коначно, интеграција додатних извора података, као што су подаци о здравственом стању животиња или подаци о условима у склоништима, могла би проширити обим и дубину анализе, омогућавајући свеобухватнији приступ управљању и планирању у склоништима за животиње.

# Литература

[1] Материјали са вежби предавања предмета „Методе и технике науке о подацима“ и „Нумерички алгоритми и нумерички софтвер“

[2] "Fuzzy Logic: Concepts, Control and Applications" – YouTube лекција