**УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ**

**ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ**

**Никола Карпић**

**Развој рјешења за предвиђање броја особа у просторији**

**дипломски рад**

**Бања Лука, фебруар 2021.**

**Тема: РАЗВОЈ РЈЕШЕЊА ЗА ПРЕДВИЂАЊЕ БРОЈА ОСОБА У ПРОСТОРИЈИ**

**Кључне ријечи:**

**Машинско учење**

**Регресија**

**Класификација**

**LightGBM**

**Комисија:**

**проф. др Милош Љубојевић, предсједник**

**проф. др Зоран Ђурић, ментор**

**Александар Келеч, ма, члан**

**Уз рад је приложен CD.**

**Кандидат:**

**Никола Карпић**

УНИВЕРЗИТЕТ У БАЊОЈ ЛУЦИ

ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ

КАТЕДРА ЗА РАЧУНАРСТВО И ИНФОРМАТИКУ

Тема: РАЗВОЈ РЈЕШЕЊА ЗА ПРЕДВИЂАЊЕ БРОЈА ОСОБА У ПРОСТОРИЈИ

Задатак: Машинско учење. Описати класификационе и регресионе алгоритме машинског учења и њихове типичне представнике. Припрема скупова података за тренирање, валидацију и тестирање. У практичном дијелу рада анализирати моделе за предвиђања броја особа у просторији креиране кориштењем неколико алгоритама (класификационих и регресионих). Моделе је потребно тренирати подацима који садрже температуру, влажност, ниво угљен-диоксида и др. За реализацију користити *Jupyter Notebook* алат. Извршити компаративну анализу перформаси добијених модела.

Ментор: проф. др Зоран Ђурић

Кандидат: Никола Карпић (1144/14)

Бања Лука, фебруар 2020.

Садржај

[1. УВОД 1](#_Toc64317729)

[2. Машинско учење 3](#_Toc64317730)

[2.1. Надгледано учење 3](#_Toc64317731)

[2.1.1. Класификација 3](#_Toc64317732)

[2.1.2. Регресија 3](#_Toc64317733)

[2.2. Ненадгледано учење 3](#_Toc64317734)

[2.3. Учење уз подстицај 3](#_Toc64317735)

[3. Процес машинског учења 4](#_Toc64317736)

[3.1. Прикупљање података 4](#_Toc64317737)

[3.2. Припрема података 4](#_Toc64317738)

[3.3. Анализа података 4](#_Toc64317739)

[3.4. Избор алгоритма 4](#_Toc64317740)

[3.5. Обучавање модела 4](#_Toc64317741)

[3.6. Оцјењивање модела 4](#_Toc64317742)

[3.7. Тестирање модела 4](#_Toc64317743)

[3.8. Примјена модела 4](#_Toc64317744)

[4. Алгоритми 5](#_Toc64317745)

[4.1. Логистичка регресија 5](#_Toc64317746)

[4.2. Gaussian Naive Byes Classifier 5](#_Toc64317747)

[4.3. K Nearest Neighbors Classifier 5](#_Toc64317748)

[4.4. Decision Tree Classifier 5](#_Toc64317749)

[4.5. Random Forest Classifier 5](#_Toc64317750)

[4.6. Gradient Boosting Classifier 5](#_Toc64317751)

[4.7. Support Vector Machine Classifier 5](#_Toc64317752)

[4.8. LightGBM Classifier 5](#_Toc64317753)

[4.9. K Nearest Neighbors Regressor 5](#_Toc64317754)

[4.10. LightGBM Regressor 5](#_Toc64317755)

[5. Практични рад 6](#_Toc64317756)

[5.1. Прикупљање података 6](#_Toc64317757)

[5.2. Припрема података 6](#_Toc64317758)

[5.3. Анализа података 6](#_Toc64317759)

[5.4. Избор алгоритма 6](#_Toc64317760)

[5.5. Обучавање модела 6](#_Toc64317761)

[5.6. Оцјењивање модела 6](#_Toc64317762)

[5.7. Тестирање модела 6](#_Toc64317763)

[6. Резултати 7](#_Toc64317764)

[7. Закључак 8](#_Toc64317765)

[8. Литература 9](#_Toc64317766)

# Увод

У данашње вријеме се суочавамо са феноменом све веће производње дигиталних података у разним форматима. Количина података се експоненцијално повећава и ти подаци су огроман ресурс који остаје неискориштен. Због наглог повећања количине података, појавили су се могућност и потреба за обрадом тих података. Како је велики проценат тих података неструктурисан и некласификован, појавила се потреба и за креирањем ефикасних алгоритама и процеса за разврставање, именовање и анализу тих података. Једна од најзаступљенијих области вјештачке интелигенције (*Artificial Intelligence*) која се бави разврставањем, именовањем и анализом података је машинско учење (*Machine learning*).

Велике компаније као што су Google[[1]](#footnote-1), Microsoft[[2]](#footnote-2), Facebook[[3]](#footnote-3) и Amazon[[4]](#footnote-4) су почеле да улажу огромне напоре и средства у своје могућности чувања, обраде и класификације великих количина несређених података како би остале релевантне и оствариле тржишну предност у данашњем брзорастућем и брзомијењајућем дигиталном пространству. Велика количина средстава која се улажу у ову област је довела до експлозије научних радова на тему машинског учења у посљедњих 15 година (слика 1.1).

Владе, војске, Министарства унутрашњих послова и остале институције разних држава улажу у ову област, такође, било због могућности да боље и ефикасније прате своје грађане и брже проналазе могуће терористе и преступнике, било због постизања предности над другим државама у војном или обавјештајном сектору.

Машинско учење је облик вјештачке интелигенције који омогућава систему да учи из података, а не путем експлицитног програмирања. Међутим, машинско учење није једноставан процес. Машинско учење користи низ алгоритама који итеративно уче из података да би побољшали, описали податке и предвидјели исходе. Како алгоритми уносе више података за учење, тако је могуће произвести све прецизније моделе засноване на тим подацима. [1]



*Слика 1.1. Број научних радова на тему машинског учења у свијету од 2000. године[[5]](#footnote-5)*

Алгоритми машинског учења се убрзано увлаче у све сфере живота данашњих људи. Све је чешће да породице посједују дигиталног асистента који се контролише звуком или помоћу паметног телефона. Све су чешћи паметни кућански уређаји и имплементације концепата као што су паметне куће са разним сензорима (камере, микрофони, детектори пожара, нивоа влаге, угљен-диоксида, освијетљености, итд...) у свим просторијама. Компаније често користе податке о броју особа у просторији, како би направиле уштеде на трошковима гријања и како би смањиле свој угљенични отисак. Смањењем угљеничног отиска компаније повећавају своју прихватљивост у очима све више еколошки освијештених потенцијалних нових клијената и генеришу милијарде долара годишње продаје. [2] Ти сензори могу да представљају проблем за осјећај приватности у дому или на радном мјесту.

Из наведених разлога се тежи да се што више корисних података добија из сензора који не нарушавају приватност директно као што то раде микрофони и камере, па се прибјегава техникама које индиректно процјењују тражене величине из података које нам дају сензори који нису толико инвазивни што се тиче приватности. Још једна брига је и очување пословних тајни у компанијама, пошто се помоћу микрофона и камера, које се могу користити за процјену броја особа у просторијама, релативно једноставно може доћи до пословних тајни изговорених на затвореним састанцима ако постоје и најмањи сигурносни пропусти у сигурносном систему компаније.

У овом раду ће се обрађивати примјењивање одређеног броја класификационих и регресионих машинских алгоритама на процјену броја особа у просторији на основу датума, времена и података прикупљених сензорима за ниво угљен-диоксида у просторији, освијетљеност просторије, количину влаге у просторији и сензорима за количину кретања у просторији. Алгоритми машинског учења које ћемо користити су: логистичка регресија, *Gaussian Naive Byes, K-Nearest Neighbors, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, LightGBM*.

У другој глави се даје објашњење вјештачке интелигенције и уопштено објашњење метода машинског учења и процеса примјене алгоритама машинског учења.

У трећој глави су детаљно објашњени кораци креирања модела машинског учења помоћу алгоритама машинског учења.

У четвртој глави су детаљније описани регресиони и класификациони алгоритми који су кориштени у овом раду.

У петој глави се налазе појединости које се односе на практични. У њој су детаљно описани подаци и примјена сваког алгоритма, те упоредна анализе резултата. Сав код је написан у програмском језику Python.

У шестој глави се налази закључак који је добијен у овом раду.

# Машинско учење

Вјештачка интелигенција је једно од најновијих поља у науци и инжењерству. Ова област је почела озбиљно да се развија убрзо након Другог свјетског рата, а само име је настало 1956. Заједно са молекуларном биологијом, вјештачка интелигенција се редовно наводи као „поље у којем бих највише волео да будем“ од научника из других дисциплина. Студент физике може оправдано да сматра да су све добре идеје већ преузели Галилео, Њутн, Ајнштајн и остали. Вјештачка интелигенција, с друге стране, још увек има мјеста за неколико нових Ајнштајна и Едисона. Вјештачка интелигенција тренутно обухвата огроман број области, у распону од општег (учење и перцепција) до специфичних, као што су играње шаха, доказивање математичких теорема, писање поезије, вожња аутомобила у препуној улици и дијагностиковање болести. Вјештачка интелигенција је релевантна за било који интелектуални задатак; то је заиста универзално поље. [3]

Као што је већ речено у уводу, машинско учење је облик вјештачке интелигенције који омогућава систему да учи из података, а не путем експлицитног програмирања. Међутим, машинско учење није једноставан процес. Машинско учење користи низ алгоритама који итеративно уче из података да би побољшали, описали податке и предвидјели исходе. Како алгоритми уносе више података за учење, тако је могуће произвести све прецизније моделе засноване на тим подацима. [1] Алгоритми машинског учења као улаз примају огромне скупове података који описују одређене појаве и обрађују их и у њима проналазе правилности, а као излаз стварају моделе који помоћу пронађених правилности могу да предвиђају резултате тих појава над невиђеним улазним подацима.

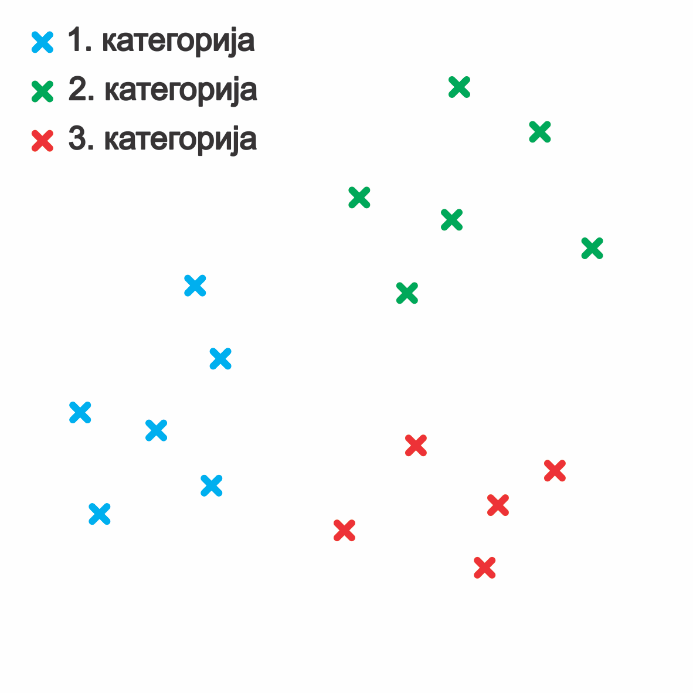
Неки од типичних примјера у којима се машинско учење показало као одговарајући алат су аутономна возила, препоручена претрага код интернетских претраживача, добијање повратние информације о томе шта купци мисле о нама, препознавање нежељених порука, откривање превара. Технике машинског учења се обично дијеле на три области у зависности од врсте повратне информације доступне систему учења, а то су: надгледано учење, ненадгледано учење и учење уз подстицај.

## Надгледано учење

Надгледано учење је врста машинског учења гдје се учи функција која пресликава улаз на излаз на основу примјера парова улаз-излаз. [3] Алгоритам учења добија скуп означених примјера као скуп података за учење и даје предвиђања за све невиђене тачке. Ово је најчешћа врста машинског учења која се користи код проблема класификације, регресије и рангирања. Проблем откривања нежељене поште је један од најпознатијих примјера код којих се користи надгледано учење. [4]

### Класификација

Класификација је проблем додељивања категорије свакој ставци из скупа података. На пример, класификација докумената састоји се од додељивања категорија као што су политика, посао, спорт или вријеме сваком документу, док се класификација слика састоји од додељивања категорије свакој слици као што су аутомобил, воз или авион. Број категорија у таквим задацима често је мањи од неколико стотина, али може бити много већи у неким тешким задацима и чак неограничен као у класификацији текста или препознавању говора. [4]



*Слика 2.1. Примјер класификације у три класе*

### Регресија

Регресија је проблем предвиђања стварне вриједности за сваку ставку из скупа података. Примери регресије укључују предвиђање вриједности залиха или варијација економских промјенљивих. У регресији, „казна“ за нетачно предвиђање зависи од величине разлике између истинске и предвиђене вриједности, за разлику од проблема класификације, гдје обично не постоји појам блискости између различитих категорија. [4]



*Слика 2.2. Примјер линеарне регресије*

## Ненадгледано учење

Ненадгледано учење је врста машинског учења код кога алгоритам учења искључиво прима необиљежене податке за учење и даје предвиђања за све невиђене тачке. Будући да уопште нема доступних означених примјера, може бити тешко квантитативно процијенити перформансе алгоритма. Груписање и смањење димензионалности су примјери проблема код којих се ненагдледано учење показало као најбоље. [4]



*Слика 2.3. Примјер смањења димензионалности ненадгледаним учењем*

### Учење уз подстицај

Учење уз подстицај је врста машинског учења код кога су фазе обуке и тестирања такође измијешане у процесу подстицања. Да би прикупио информације, алгоритам учења активно комуницира са окружењем и, у неким случајевима, утиче на окружењем и за сваку радњи прима тренутну „награду“ у облику неке нумеричке вриједности. Циљ алгоритма учења је максимизирање његове награде током процеса подстицања. Међутим, окружење не пружа повратне информације о дугорочним наградама, а алгоритам учења се суочава са дилемом истраживање или експлоатација, јер мора да бира између истраживања непознатих радњи (како би стекао више информација) и искориштавања већ прикупљених информација. [4]



*Слика 2.4. Илустрација агента који учи уз подстицаје*

# Процес машинског учења

## Прикупљање података

## Припрема података

## Анализа података

## Избор алгоритма

## Обучавање модела

## Оцјењивање модела

## Тестирање модела

## Примјена модела

# Алгоритми

## Логистичка регресија

## Gaussian Naive Byes Classifier

## K Nearest Neighbors Classifier

## Decision Tree Classifier

## Random Forest Classifier

## Gradient Boosting Classifier

## Support Vector Machine Classifier

## LightGBM Classifier

## K Nearest Neighbors Regressor

## LightGBM Regressor

# Практични рад

## Прикупљање података

## Припрема података

## Анализа података

## Избор алгоритма

## Обучавање модела

## Оцјењивање модела

## Тестирање модела

# Резултати

# Закључак

# Литература

|  |  |
| --- | --- |
| [1] | D. Kirsch и J. Hurwitz, Machine Learning For Dummies, New Jersey: John Wiley & Sons, Inc., 2018. |
| [2] | F. Williams, Meet the nine billion-dollar companies turning a profit from sustainability, The Guardian, 2016. |
| [3] | T. Mitchell, Machine Learning, New York: McGraw-Hill, 1997. |

1. <https://www.google.com/> [↑](#footnote-ref-1)
2. <https://www.microsoft.com/> [↑](#footnote-ref-2)
3. <https://www.facebook.com/> [↑](#footnote-ref-3)
4. <https://www.amazon.com/> [↑](#footnote-ref-4)
5. <https://www.forbes.com/sites/louiscolumbus/2018/01/12/10-charts-that-will-change-your-perspective-on-artificial-intelligences-growth> [↑](#footnote-ref-5)