УНИВЕРЗИТЕТ У БЕОГРАДУ

ФАКУЛТЕТ БЕЗБЕДНОСТИ
Катедра студија безбедности

Примена машинског учења и биометрије понашања на безбедност веб апликација

– ДИПЛОМСКИ РАД –

Ментор:
Ана Ковачевић
Проф. др

Студент:
Невена Дабовић
123/10

Београд, 2022.

САДРЖАЈ

1. Увод

Свет у коме данас живимо је монетизован и дигитализован, уз складиштење велике количине података. Складиштење података, нарочито у дигиталном свету, са собом носи повећан ризик од превара, крађа или других малициозних упада на приватне и корпоративне системе. Циљ таквих упада је скоро увек крађа информација, било да су то информације о корисницима, корисничким налозима, о приступу банковним рачунима, програмерском коду или нечем другом.

Компанија A10 Networks је објавила списак од 10 предикција за сајбер безбедност (A10 Staff, 2016) у непосредној будућности:

1. Дигитална безбедност ће постати основно питање људских права;
2. Хакерски напади могу осакатити, делимично или потпуно, главне оператере мобилних мрежа циљајући њихова језгра;
3. Енкрипција ће постати много важнија у онлајн саобраћају, нарочито на глобалним релацијама између истока и запада;
4. Локалне и централне владе ће доживети више сајбер напада него икада раније;
5. Безбедносне и аналитичке функционалности без коришћења сервера ће постати популарније за коришћење функција попут скенирања вируса;
6. Провајдери клауд система постају главна мета нападача;
7. Прилагодљиви и реакциони безбедносни производи ће се котирати у пет најбољих технологија;
8. Вештачка интелигенција ће се у великој мери користити за покретање нових безбедносних технологија;
9. Рањиви (незаштићени) системи контроле и прикупљања података ће проузроковати физичку штету у непосредној будућности;
10. Бићемо сведоци успона блокчејн (blockchain) безбедносних технологија.

Већина ових предикција су се обистиниле, па није изненађујуће што су модели безбедности побољшани и неке нове технологије почеле да се развијају. Те технологије (које нису толико нове, колико имају нову, безбедносну примену) су везане за коришћење вештачке интелигенције и њених области, нарочито машинског учења.

По истраживањима великог безбедносног бренда Trustwave, 100% сајбер напада на веб апликације долази из проблема аутентификације (Trustwave, 2018), те машинско учење почиње да интегрише специфичан скуп података - у науци познатији као биометрија. Скуп биометријских података који се користе приликом примене машинског учења се назива биометријом понашања. Пошто многи компјутерски системи немају прилику да потврде идентитет корисника у реалном времену користећи податке физичке биометрије (попут скенирања ока или отиска прста), користи се другачија врста података који потпадају под појам биометрије понашања.

Овај рад представља преглед основних принципа и подела биометрије, машинског учења и начина на који те две научне гране корелирају у пракси, конкретно приликом обезбеђивања приступа веб апликацијама.

2. Биометрија понашања

Биометрија је аутоматизовано распознавање појединаца помоћу јединствених физичких и психичких особина. Неретко се среће дефинција која биометрију карактерише као скуп телесних мерења и прорачуна везаних за људско тело.

Термин "биометрија" изведен је од грчких речи "биос" (живот) и "метрео" (мерити). Модерни, аутоматизовани биометријски системи постали су доступни тек у последњих неколико деценија, вођени значајним успехом у области истраживња рачунарске обраде података. Многе од ових нових аутоматизованих техника су, међутим, засноване на идејама које су испрва зачете вековима раније.

2.1. Историјски развој биометрије

Један од најстаријих и најосновнијих примера физиче карактеристике коју људи користе за препознавање је лице. Од почетка цивилизације, људи су користили лица да идентификују познате и непознате појединце из околине. Таква идентификација појединаца постајала је све већи изазов како се популација повећавала и како су бољи путеви и погоднија превозна средства увели многе нове појединце у некада мале заједнице (Mayhew, 2018).

Концепт препознавања од човека до човека такође се може базирати на другим особинама - попут гласа и става (начина на који се човек носи, хода, гестикулира). Људи подсвесно користе ове карактеристике да издвоје познате појединце свакодневно.

Историјски гледано, коришћени примери распознавања су се развијали експоненцијално. Занимљиве су тезе Стивена Мејхјуа које систематизују примере распознавања од коришћених свакодневно у друштву, до хијерархијски и научно утемељених (Mayhew, 2018):

- У пећини у Француској (процењене старости око 31.000 година) зидови су украшени сликама за које се верује да су их створили праисторијски људи

који су тамо живели. Око ових слика налазе се бројни отисци руку за које се сматра да су служили као потпис.

- Такође, постоје докази да су отисци прстију коришћени као знак распознавања особе још 500. године пре нове ере. Вавилонске пословне трансакције су забележене у глињеним плочама које садрже отиске прстију.
- Жоао де Барос, шпански истраживач и писац, написао је да су рани кинески трговци користили отиске прстију за поравнање пословних трансакција. Кинески родитељи су такође користили отиске прстију и стопала да би разликовали децу једно од другог.
- У раној египатској историји, трговци су идентификовани према њиховим физичким карактеристикама да би се разликовали између поверљивих трговаца и претходних успешних трансакција, и оних који су нови на тржишту.
- Персијска књига из 14. века „Jaamehol-Tawarikh“садржи коментаре о пракси идентификације особа по отисцима прстију.
- Године 1684. Др Нехемија Гр је у „Филозофским трансакцијама Лондонског краљевског друштва” говорио о различитим испупченим деловима на кожи који су јединствени за појединце.
- Књига Говарда Бидлуа из 1685. године, „Анатомија људског тела“, такође описује детаље коже на превојима.
- Године 1686. Марчело Малпиги, професор анатомије на Универзитету у Болоњи, описао је у својим делима отиске прстију, спирале и петље.

До средине 1800. година, са брзим растом градова због индустријске револуције и продуктивније пољопривреде, постојала је формално призната потреба за идентификацијом људи (Mayhew, 2018). Трговци и власти су се суочавали са све гушћом популацијом и више се нису могли ослањати само на сопствена искуства и познанике. Под утицајем списа Џеремија Бентама и других утилитаристичких мислилаца, правни системи овог периода почели су да кодификују концепте

правде који су присутни код нас до данас (Mayhew, 2018). Најважније, правосудни системи су настојали да блаже поступају према онима који су први пут прекршили закон, а према повратницима строже. Ово је створило потребу за формалним системом који би евидентирао прекршаје заједно са описима физичких особина преступника.

Први од два приступа био је Бертијонов систем мерења различитих телесних димензија, који је настао у Француској. Ова мерења су записана на картицама које су се могле сортирати по висини, дужини руке или било ком другом параметру (Mayhew, 2018). Ова област се звала антропометрија.

Други приступ је био формална употреба отисака прстију од стране полицијских управа. Овај процес се појавио у Јужној Америци, Азији и Европи. До касних 1800. година развијена је метода за индексирање отисака прстију која је пружала могућност преузимања записа као што је то урадила Бертијонова метода, али се заснивала на индивидуализованијим метричким обрасцима отисака прстију и шаке. Први такав свеобухватан систем за индексирање отисака прстију развио је у Индији Азизул Хаке (Azizul Haque) за Едварда Хенрија, генералног инспектора полиције у Бенгалу. Овај систем, назван Хенријев систем, и његове варијације се још увек користе за класификацију отисака прстију (Souza & Medeiros, 2022).

Биометријски системи у форми какву данас познајемо, почели су да се појављују у другој половини двадесетог века, што се подудара са појавом компјутерских система. Поље у настајању доживело је експлозију активности током деведесетих и почело је да се појављује у свакодневним применама раних двехиљадитих.

2.2. Подела биометрије по врстама

Основна подела биометрије је на физиолошку (у различитој литератури се може наћи и под именом физичка) и биометрију понашања (бихејвијоралну).

Две дефиниције Ванг и Генга представљају физиолошку биометрију као скуп података везаних за специфичне мере, димензије и карактеристике људског тела. Биометрија понашања уочава обрасце понашања индивидуе и такви подаци се могу прикупљати не само једном већ током читавог анализираног периода, укључујући и реално време (Wang & Geng, 2009).

По њиховом истраживању, физиолошка биометрија се дели на (Wang & Geng, 2009):

1. **Биометрију уха**

   Људско ухо има специфичан, јединствен облик и карактеристике које омогућавају идентификацију појединца. Идентификација ушију се користи дуги низ година у земљама попут Француске, а детаљи о ушима ухапшене особе прикупљани су заједно са сликама лица и отисцима прстију као део кривичног досијеа. Са појавом напредних рачунарских алгоритама као што су конволуционе неуронске мреже (Convolutional Neural Networks), биометрија уха је постала део одрживе аутоматизоване биометријске базе која сада превазилази своје традиционалне примене (попут законске).

2. **Биометрију очних вена**

   За биометријску идентификацију очију користи се узорак формиран од вена беоњаче. Људске очи, са својим различитим структурним елементима, пружају могућност личне идентификације. Препознавање очних вена функционише помоћу шаблона шаренице у људском оку. Шареница је обојено ткиво које има кружни отвор у средини. Сваки појединац има различите шаре и боје шаренице, па чак је и шара шаренице левог и десног ока различита. Ова технологија је могућа чак и на даљину, па је стога

погодна за многе примене. Ову технологију је развио др Реза Деракшани, професор рачунарства и електротехнике на Универзитету Мисури.

3. Биометрију лица

Биометрија лица користи фацијалне карактеристике за идентификацију појединца. Постоји широк спектар техника које се користе за статистичко сагледавање карактеристика лица на начин на који старост, израз лица, осветљење или многе друге варијабле значајно утичу. Такве технике могу укључивати алгоритме машинског учења, као што су конволуционе неуронске мреже, које су тренирање на огромним скуповима слика лица које не укључују директно мерење удаљености између карактеристика. Тренутни алгоритми лица описују облик и изглед црта лица, као што су очи, нос или уста, применом обраде слике која је посебно обучена да обухвати промењиве и статичне податке комбиноване у нумеричку репрезентацију познатију као шаблон лица.

4. **Биометрију вена прстију**

Препознавање вена прстију је технологија препознавања која ради на јединственом узорку крвних судова који се налази испод омотача коже људског прста. Овај образац формирају крвни судови и вене које носе крв према срцу. Јапанска компанија Хитачи је представила идентификацију вена (vein ID), технологију аутентификације засновану на узорку вена људског прста, у свом годишњем извештају иновација (Hitachi, 2005). Не постоји јединствен начин на који ће крвни судови појединца пролазити кроз прст (или длан, или на другим местима), те се значајно разликују међу појединцима. Овај јединствени образац формиран од вена на људском прсту ускоро ће се интензивно користити за личну идентификацију и аутентификацију.

5. Биометрију отиска прста

Отисци прстију се формирају од издигнутих папиларних гребена који пролазе преко површине коже. Људи, заједно са неким другим сисарима, имају ове гребене на прстима, палчевима, длановима, прстима и табанима. Гребени су еволуирали да обезбеде трење како би помогли хватање и кретање. Ток ових гребена често формира шаблоне, али сами гребени имају одступања у њиховој структури. Појава ових детаља је по природи насумична и користе се као основа за утврђивање идентитета јер ниједна два региона коже, која носе систем папиларних гребена, никада нису имала исти шаблон. Сходно томе, отисци прстију на свакој људској јединки су јединствени и могу се користити за идентификацију појединаца. Исто важи и за отиске дланова, али површина избочене коже је много већа и стога садржи више детаља. Поједини биометријски системи користе отиске дланова заједно са отисцима прстију за досијее појединаца.

6. ДНК анализу

ДНК (дезоксирибонуклеинска киселина) је хемијска супстанца која се налази у свакој од око 100 трилиона ћелија у људском телу. Садржи информациони, генетски код за умножавање ћелија и конструисање протеина потребних за одржавање и развој живота. Целокупна ДНК у свакој ћелији садржи комплетан сет биолошких упутстава за стварање организма и позната је као геном. ДНК пронађена у језгру ћелије подељена је на два хромозома (један наслеђен од мајке, а други од оца) и овај ДНК материјал садржи и регионе који кодирају протеине и регионе који не кодирају. Регион који кодира протеин је познат као ген и садржи све информације за ћелију како би производила протеине. Гени формирају мање од 5% генома који се углавном састоји од некодирајуће ДНК.

Значајно се разликују базе са ДНК подацима које се користе у медицинским истраживањима или генеологији, од оних које се тичу протокола за обраду идентитета и биометријских база података. ДНК је присутна у целом

људском ћелијском материјалу (коса, крв, кожа итд.), али биометријски узорак се обично узима узимањем бриса пљувачке са унутрашње површине образа да би се уклониле ћелије коже. Стога, биометријско узорковање ДНК захтева контакт са субјектом.

7. **Биометрију отиска и динамике стопала**

Попут отисака прстију и длана, људски отисак стопала се такође сматра јединственом физичком особином и омогућава идентификацију јединке. Структура гребена људског стопала остаје иста током целог живота особе, као што се избочине на длану и отисци прстију не мењају за цео живот. Дакле, то нам даје прилику да користимо отиске стопала као биометријско средство. Основна технологија за скенирање и обраду отисака остаје мање-више иста као и друга технологија за препознавање гребена прста. Технологија скенирања отиска стопала није још увек развијена у целости, те се и даље експериментише са различитим приступима.

8. **Биометрију хода**

Сваки човек има специфичан начин ходања и трчања. Фактори као што су укупна грађа особе, дужина и ширина корака, брзина кретања, различити углови формирани од зглобова кука, колена и скочног зглоба, као и углови трупа, бутина и стопала могу се анализирати и биометријски категоризовати. Према томе, појединци се могу идентификовати на основу њиховог стила хода и то омогућава стопроцентну биометријску верификацију, али сужава идентификацију на 1 према X (као могућност претраживања). Зато се користи као додатак другим биометријским анализама а не изоловано, и највећу примену има у медицинским истраживањима и дијагностичком тестирању, спортској науци и медицини, итд.

Биометрија понашања се дели на:

1. Покрете приликом коришћења уређаја (машине),
2. покрете тела и
3. гласовне варијације.

2.3. Биометрија понашања

Биометрија понашања идентификује обрасце у начинима на које људи обављају одређене задатке – обрасци у ходању, говору, куцању на тастатури или чак померању курсора (миша). Ове обрасце је изузетно тешко забележити и поновити и они се временом развијају (Wang & Geng, 2009).

Бихејвиорални биометријски алати анализирају ове обрасце, а затим се адаптирају развоју корисника. Они користе развијене статистичке моделе и машинско учење како би уочили разлике између постепеног еволуирања познатог корисника и нежељеног присуства потпуно другог корисника.

Као што је наглашено у претходном поглављу, биометрија понашања се дели на три типа, која се у пракси често користе симултано (или комбиновано) кроз систем машинског учења.

2.3.1. Покрети приликом коришћења уређаја

Начин на који особа користи свој уређај (мобилни телефон, лаптоп, стони компјутер, таблет, итд) оставља својеврсни дигитални отисак. Могу се идентификовати следеће области за анализу података:

- динамика притискања тастера (или дугмади на тастатури),
- мобилне интеракције и
- кретање курсора.

2.3.2. Динамика притискања тастера

Динамика притискања тастера или биометрија куцања на тастатури, односи се на аутоматизовани метод идентификације или потврђивања идентитета појединца на основу начина и ритма куцања на тастатури. Већ током Другог светског рата војна обавештајна служба је користила технику познату као "песница пошиљаоца

(The Fist of the Sender)" да на основу ритма разликује да ли је поруку Морзеовим кодом послао савезник или непријатељ. (Kochegurova, Gorokhova & Mozgaleva, 2017). Данас, свако домаћинство има најмање једну компјутерску тастатуру, што динамику притиска на тастере чини најлакшим, биометријским, хардверским, двофакторским аутентификацијским решењем.

Са динамиком притиска на тастере, биометријски шаблон који се користи за идентификацију појединца заснива се на обрасцу куцања, ритму и брзини куцања на тастатури. Необрађена мерења која се користе за динамику притиска на тастер су "време задржавања (dwell time)" и "време преласка (flight time)":

- Време задржавања је време трајања притиска на тастер;
- Време преласка је време које траје између отпуштања тастера и притиска на следећи тастер.

Када се куца низ знакова, време које особи треба да пронађе прави тастер (време преласка) и време док држи тастер (време задржавања) су специфична за тај субјект и могу се израчунати независно од укупне брзине куцања. Ритам којим се куцају неки низови знакова веома зависи од особе. На пример, неко ко је навикао да куца на енглеском језику ће брже куцати одређене секвенце знакова као што је „the"него особа која је навикла на француски језик (Monrose & Rubin, 2000).

Многи софтвери комбинују динамику притиска на тастере са другим интеракцијама које корисник има са рачунаром, као што су покрети миша (време убрзања, учесталост кликова).

Када се динамика притиска на тастер користити за аутентификацију, углавном се користи заједно са корисничком идентификацијом (именом) и лозинком као облик двофакторске аутентификације.

Друга употреба је као врло специфичан облик надзора. Постоје софтверска решења која, (без знања крајњих корисника) прате динамику притиска на тастере за сваки кориснички налог. Ово праћење и архивирање динамике притиска на тастере се затим користи за анализу да ли налоге деле или их уопште користе

други људи, поред стварног власника налога. Разлози за такву имплементацију могу бити верификација корисника (коју прате безбедносне процедуре) или провера да се не деле софтверске лиценце (посебно за Software as a service - SaaS апликације).

Уопштено говорећи, биометрија понашања, као што је динамика притиска на тастер, мање је поуздана од физиолошке биометрије те се користи у комбинацији са другим биометријским типовима. У пракси се користи 7 критеријума за процену прикладности и примењивости динамике притиска на тастере - универзалност, јединственост, трајност, доступност прикупљивости података, прихватљивост, заобилажење законских регулатива и перформансе корисника, и ови критеријуми,систематизовани од стране Монроса и Рубина се користе у пракси већ 22 године (Monrose & Rubin, 2000).

Технике које се користе за динамику притиска на тастере веома се разликују по снази и софистицираности, од статистичких приступа до неуронских мрежа. Неколико области које се мере приликом креирања шаблона за динамику притиска на тастере укључују:

1. Сирову брзину куцања (без исправљања грешака)
2. Време тражења и време чекања
3. Карактеристике словних секвенци
4. Карактеристике уобичајених грешака
5. Замене, преокрете и испадања из шаблона куцања

Највећи проблем код прикупљања и систематизације ових података јесте људски фактор (неретко је случај да људи у исто време куцају на тастатури и раде још нешто, на пример једу, па подаци неће бити конзистентни и релевантни).

2.3.3. Мобилне интеракције

Данас тржиште нуди хиљаде мобилних уређаја са сензорима који се могу користити за биометријско препознавање. Модерни паметни телефони имају камере, микрофоне и екране осетљиве на додир. Ово омогућава брзу

аутентификацију путем лица, гласа или помераја прстију на екрану. Са развојем мобилне технологије и потражње тржишта, тако је биометрија постајала широко распрострањена.

Мобилне интеракције обухватају све што можете да урадите са екраном осетљивим на додир: превлачење, тапкање, притисак, куцање или зумирање (увеличавање) прстима. Приликом анализе употребе курсора узимају се у обзир брзина, кликови, путање и промене правца.

Користе се исти принципи гестикулације, динамике притискања тастера, препознавања гласа и кретања курсора као код осталих принципа биометрије понашања, са том разликом што се углавном користе комбиновано (multimodal authentication).

Због непоузданих карактеристика које би могле да буду узроковане једном биометријском карактеристиком (тј. променом емоционалног или физичког стања корисника или лошим прикупљањем података), и да би се превазишла деградација перформанси изазвана овим ограничењима, истраживачи су прешли са унимодалне биометрије на мултимодалну биометрију (Baltrušaitis, Ahuja & Morency, 2018). На пример, комбиновање препознавања лица и динамике притиска на тастер за аутентификацију корисника ће надмашити перформансе сваког самосталног модалитета и ублажити ограничења сваког модалитета појединачно (Baltrušaitis, Ahuja & Morency, 2018).

Међутим, методе засноване на мултимодалним методама се суочавају са неколико изазова, јер док користе неколико модалитета за повећање перформанси, алгоритам за аутентификацију постаје оптерећен јер треба да научи образац корисника у различитим модалитетима. Да би превазишли овај проблем, истраживачи користе скуп модела машинског учења који омогућавају различито препознавање образаца по легитимном кориснику (Baltrušaitis, Ahuja & Morency, 2018).

Ефикасно решавање таквих изазова омогућава мултимодалној аутентификацији да понуди свеобухватну и безбедну контролу приступа.

Имплементација мултимодалне аутентификације захтева синтезу више извора података, инстанцираних карактеристика или/и коришћених алгоритама и модела. Истраживања показују да мултимодалне биометријске шеме аутентификације користе различите синтезе модела као што су синтеза карактеристика, синтеза коришћених алгоритама за моделирање и синтеза на нивоу одлуке (Baltrušaitis, Ahuja & Morency, 2018).

2.3.4. Кретање курсора

Биометријски систем кретања миша састоји се од три компоненте (Baltrušaitis, Ahuja & Morency, 2018):

- прикупљање података о мишу,
- издвајање карактеристика и
- класификација образаца.

Када корисник почне да користи миш за обављање својих задатака, програм за праћење који ради у позадини прикупља податке. Датотека необрађених података који се прикупљају садрже следеће информације за сваки унос:

• Догађај миша (mouse event), било да се ради о померању, превлачењу или клику;

• Време догађаја у милисекундама;

• Координате курсора на корисничком екрану.

Након прикупљања података, издвајају се карактеристике које служе за класификацију образаца понашања. Из тих мерења креирамо вектор карактеристика, који заузврат представља кориснички профил или потпис корисника.

Необрађени подаци прикупљени из модула за интелигентно прикупљање података (data mining/data collection module) се обрађују да би се креирале криве кретања миша и кликови мишем. Свака крива кретања и клик су повезани

величином криве, дужином криве, брзином криве, убрзањем криве, трајањем клика и кривином криве (Baltrušaitis, Ahuja & Morency, 2018).

На крају поступка, класификују се обрасци добијени комбинацијом информација о криви кретања миша и кликова. Резултат класификације су јасни кориснички профили, односно потпис корисника. У бази података, тај потпис може да се упореди са постојећим корисницима и да се јасно спари са посматраним корисником.

2.3.5. Покрети тела

Подаци о људском телу се односе на кинезиолошке податке, односно став, постуру или држање тела.

Кинезиологија уочава да се као став тела, тачније, релативна усклађеност делова тела једног у односу на други, при чему је напрезање тела најмање. Човека карактерише усправан стојећи став (Simeón & Monge, 2005).

Тело човека је симетрично. Тело се, зарад процене и анализе, уздужно дели на леву и десну идентичну половину. Замишљена раван која дели тело уздужно почев од чела, преко корена и врха носа, средине доње вилице, грудне кости, пупка, пубичне кости, све до простора на поду између оба стопала, зове се сагитална раван (Simeón & Monge, 2005).

Тело се, зарад процене и анализе, дели и попречно на предњу (лице, предњи део грудног коша и предњи део ногу) и задњу страну (потиљак, задњи део грудног коша-леђа, глутеална регија и задњи део ногу). Замишљена раван полази од врха рамена, ушне шкољке, бочне средине тела и зове се фронтална раван (Simeón & Monge, 2005).

Тело се зарад процене и анализе дели и хоризонтално на горњу и доњу половину. Горњу половину чине труп и горњи екстремитети. Доњу половину тела чине карлица и доњи екстремитети.Замишљена раван полази изнад карлице,

хоризонтално и зове се хоризонтална раван. Сви покрети тела се изводе у ове три равни (Simeón & Monge, 2005).

Одступања од симетричности представљају поремећаје у постури и биомеханици. Свако држање тела даје јединствен биометријски отисак.

Држање тела се посматра двосмерно:

- Динамично држање је начин на који се особа држи док се креће, на пример када хода, трчи или се сагиње да би нешто подигла.
- Статички положај је начин на који се особа држи када се не креће, на пример када седи, стоји или спава.

- Начин хода

Препознавање хода се заснива на идеји да свака особа има карактеристичан и јединствен начин хода, који се лако може уочити са биомеханичке тачке гледишта. Људски покрет се састоји од синхронизованих покрета стотина мишића и зглобова, иако су основни обрасци покрета слични, ход се разликује од особе до особе у смислу времена и величине. Као последица тога, мање варијације у стилу хода могу се користити као биометријски идентификатор за идентификацију појединаца (Cao, 2020).

Препознавање хода групише просторно-временске параметре, као што су дужина корака, ширина корака, брзина хода и време циклуса са кинезиолошким параметрима, као што су зглобна ротација кука, колена и скочног зглоба, средњи зглобни углови кука, колена и скочног зглоба и бутине, углови трупа и стопала. Такође се разматра корелација између дужине корака и висине појединца (Cao, 2020).

Пошто је људско кретање један облик људских покрета, препознавање хода је уско повезано са методама вида које откривају, прате и анализирају људско понашање у анализи људских покрета. Технологије за препознавање хода су тренутно у повоју. Тренутно постоје две главне врсте техника за препознавање хода (Cao, 2020):

1. Први је препознавање хода на основу аутоматске анализе видео снимака. Овај приступ је најпопуларнији приступ проучаван и укључује анализу видео узорака хода субјекта и путање зглобова и углова. Прави се математички модел кретања, који се затим упоређује са другим узорцима како би се утврдио идентитет.
2. Друга метода користи радарски систем, који бележи циклус хода који стварају различити делови тела субјекта. Ови подаци се затим упоређују са другим узорцима како би се извршила идентификација.

У оба модела, анализа људског тела је примењена на ненаметљив начин користећи техничке инструменте који мере покрете тела, телесну механику и активност одређених мишићних група. Такве технологије су пројектоване за употребу у кривичном правосуђу и националној безбедности.

- Начин руковања предметима

2.3.6. Гласовне варијације

Глас особе, тј. начин на који звучи када говори је резултат комбинације специфичних физичких атрибута (као што су дужина гласних жица и облик грла) и специфичних атрибута понашања (као што је акценат са којим особа говори).

Људски глас ствара таласне дужине које се могу измерити. Глас се прикупља и анализира помоћу софтвера који користи вештачку интелигенцију и технике машинског учења како би произвео широку лепезу података изведених из фактора као што су модулација говора, тонови, нагласак, фреквенција итд (Markowitz, 2000). Ови елементи омогућавају систему да креира референтни шаблон за глас (познат као 'гласовни отисак' или 'гласовни модел) који се може користити за аутентификацију говорника у наредним трансакцијама (накох прве, када се оставља поменути отисак) (Markowitz, 2000). Слична технологија се примењује како би се омогућило уређајима да разумеју, преводе и комуницирају са гласовном командом/питањем, на пример, када разговарају са паметним

звучницима, мобилним уређајима, кућним апаратима, виртуелним помоћницима.

Напомена: Постоји разлика између препознавања говорника (препознавање ко говори) у биометријским апликацијама и препознавања говора (препознавање онога што се говори) нпр. апликације као што су машински диктат, системи гласовних команди, итд препознају оно ШТО се говори. Ова два појма се често мешају. Једноставно речено, глас је синоним за говорника, а не за говор.

Препознавање гласа се може користити и за стопроцентну верификацију и за биометријске начине идентификације 1:X (један према више - када се идентификује један, непознати глас међу више корисника) (Markowitz, 2000).

У режиму стопроцентне верификације, препознавање говорника користи глас као метод аутентикације идентитета говорника. Често се користи као „чувар капије“(gatekeeper) да би се обезбедио приступ безбедном систему (телефонско банкарство као пример примене). Корисник је свестан рада овог система и обично се захтева сарадња између корисника и система.

Процес регистрације и верификације гласа и препознавање гласа могу да функционишу у било ком окружењу где нема претеране буке. Посебно је ефикасан приликом електронског комуницирања. Због тога је примењив на низу различитих случајева у којима идентитет говорника треба да буде аутентификован како би им се омогућило да траже услуге, предузимају трансакције, издају команде или снимају сложене вербалне информације.

Малициозни напади у системима за препознавање гласа су важан фактор и они укључују изазове као што су имитација гласа, синтетичка конверзија гласа софтвером и снимљени па реплицирани глас. Мере против лажирања укључују откривање „живости“током трансакције и присуство различитих лажних артефаката у гласовним отисцима (Markowitz, 2000).

3. Машинско учење

Машинско учење је област вештачке интелигенције и информационих наука која се фокусира на коришћење података и алгоритама за имитирање начина на који људи уче, постепено побољшавајући његову тачност (Jordan & Mitchell 2015).

Артур Семјуел је сковао термин „машинско учење“приликом свог истраживања базираног на игри Даме (Checkers). 1962. године, Роберт Нили, самопроглашени мајстор Даме, губи партију против компјутера IBM 7094. Иако наизглед тривијалан за данашње стандарде, овај подвиг се сматра великом прекретницом у области вештачке интелигенције (Samuel, 1967). Током наредних неколико деценија, технолошки развој, нарочито у доменима складиштења података и снаге процесора, омогућио је иновативност на пољу машинског учења (на пример, у пракси примењено на Нетфликсов систем препорука или самоуправљајући аутомобил).

Машинско учење је интергална компонента развијајућих наука које користе информационе базе података. Коришћењем статистичких метода, алгоритми машинског учења се тренирају да податке класификују и праве предвиђања, правећи основу за пројекте интелигентног тражења или интелигентног прикупљања података. Овај систем касније олакшава доношење одлука унутар апликација, али и генерално у фирмама, предузећима и корпорацијама, утичући на кључне показатеље раста. “Big data”, односно, велике скупине података све више расту унутар система правних лица, стога и сама потражња за квалификованим особљем у овој сфери расте. Њихов главни задатак постаје да помогну у идентификацији најрелевантнијих пословних питања, а затим и података за одговоре на њих.

3.1. Вештачка интелигенција

Термин “вештачко” се односи на нешто што је направио човек, а “интелигенција” значи способност разумевања или размишљања. Погрешно је схватање да је

вештачка интелигенција систем. Вештачка интелигенција је имплементирана унутар система (McCarthy, 2007). Постоји много дефиниција вештачке интелигенције, те је једна од њих - „Студија о томе како обучити рачунаре тако да могу да раде ствари које тренутно људи могу да раде боље.“ Стога је то интелигенција којој се на људске способности, додају и машинске.

Џон Макарти нуди следећу дефиницију у чланку о појму вештачке интелигенције: „То је наука и инжењеринг прављења интелигентних машина, посебно интелигентних компјутерских програма. Повезано је са коришћењем рачунара за разумевање људске интелигенције, али не мора да се ограничи на методе које су биолошки видљиве (McCarthy, 2007)."

Међутим, деценијама пре ове дефиниције, рођење разговора о вештачкој интелигенцији означено је кључним делом Алана Тјуринга, „Рачунарска машина и интелигенција“, које је објављено 1950. године. У раду, Тјуринг, који се често назива „оцем компјутерске науке“, поставља следеће питање: „Могу ли машине да мисле?“ Одатле, он нуди тест, сада познат као "Тјурингов тест", где би људски испитивач покушао да направи разлику између компјутерског и људског текстуалног одговора. Иако је овај принцип био много пута тестиран и оспораван, он остаје важан део историје вештачке интелигенције и филозофског концепта (Turing, 1950).

3.1.1. Циљеви вештачке интелигенције

Један од водећих уџбеника у проучавању вештачке интелигенције написан од стране Стјуарт Расела и Питер Норвига је “Вештачка интелигенција: модеран приступ”. У њему дефинишу четири потенцијална циља вештачке интелигенције, који разликују рачунарске системе на основу рационалности и размишљања насупрот имитације интелигенције (Norvig & Russell, 2020):

Људски приступ:

- Системи који мисле као људи;
- Системи који се понашају као људи.

Идеалан приступ:

- Системи који размишљају рационално;
- Системи који делују рационално.

Дефиниција Алана Туринга би потпала под категорију „система који се понашају као људи“.

У свом најједноставнијем облику, вештачка интелигенција је поље које комбинује рачунарску науку и свеобухватне скупове података, како би се омогућило решавање проблема. Такође обухвата подобласти машинског учења и дубоког учења (deep learning). Ове дисциплине се састоје од алгоритама вештачке интелигенције који настоје да створе високо специјализоване системе који праве предвиђања или класификације на основу улазних података.

3.1.2. Подела вештачке интелигенције по типовима

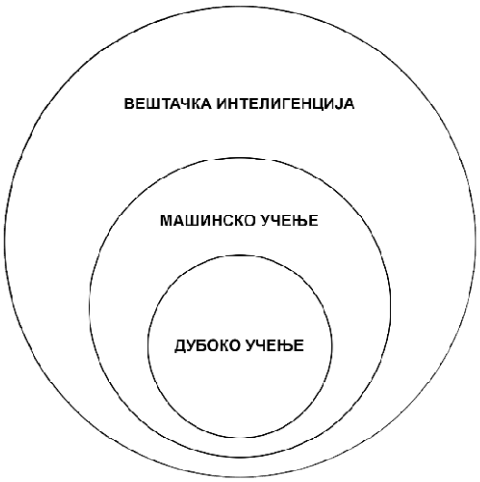
Основна подела вештачке интелигенције је на слабу и јаку (Salekar, 2022):

- Слаба вештачка интелигенција—која се назива и уска вештачка интелигенција - је вештачка интелигенција истренирана и фокусирана на обављање одређених задатака. Слаба вештачка интелигенција чини већину вештачке интелигенције која нас данас окружује. „Уско“би могло бити тачнији опис ове врсте вештачке интелигенције, јер је све само не слаба - омогућава рад неке веома широко распрострањених апликација, као што су Еплов Сири, Амазонова Алекса, ИБМ-ов Вотсон и самоуправљајућа возила.
- Јака вештачка интелигенција се састоји од вештачке опште интелигенције и вештачке супер интелигенције. Вештачка општа интелигенција, је теоријски облик вештачке интелигенције где би машина имала интелигенцију једнаку људима; имала би аутономну свест која има способност да решава проблеме, учи и планира будућност. Вештачка супер интелигенција (суперинтелигенција) премашила би интелигенцију и способност људског мозга. Јака вештачка интелигенција још увек потпуно теоретска без практичних примера у употреби.

3.1.3. Области вештачке интелигенције

Два велика научна поља припадају вештачкој интелигенцији као њене области - машинско учење и дубоко учење.

Пошто се дубоко учење и машинско учење обично користе наизменично, морају се напоменути нијансе између њих. Као што је горе поменуто, и дубоко учење и машинско учење су подобласти вештачке интелигенције, а дубоко учење је заправо подобласт машинског учења.



Слика 1: Области вештачке интелигенције

Дубоко учење је класа алгоритама машинског учења који (Introduction to Deep Learning, 2019):

- користе вишеслојне, нелинеарне процесорске јединице за извлачење и трансформацију карактеристика. Сваки следећи слој узима као улаз излазне елементе претходног слоја;
- уче на надгледан и/или ненадгледан начин;
- уче више начина презентовања података, на више нивоа који одговарају различитим степенима апстракције;
- користе облик алгоритма са опадајућим градијентом за тренинг кроз повратно пропагирање грешке (feedback learning);

- Слојеви коришћени у дубоком програмирању укључују скривене слојеве вештачке неуронске мреже и мноштво одговарајућих формула. Могу укључити и слојевито организоване скривене променљиве.

Класично, или „не-дубоко“, машинско учење више зависи од људске интервенције у учењу. Стручњаци одређују хијерархију карактеристика да би разумели разлике између уноса података, што обично захтева више структурираних података за учење.

3.2. Шта је машинско учење?

Машинско учење је област проучавања која рачунарима даје могућност да уче без експлицитног програмирања. Машинско учење користи огромну количину структурираних података тако да може да генерише резултате или даје предвиђања на основу датих података. Тренутно се машинско учење користи у алгоритмима Гугл претраге, филтерима за нежељену пошту, за предлоге Фејсбук пријатеља и препорукама за интернет куповину.

Приступи машинског учења помажу у идентификацији, функцијама препознавања као и класификацији која је неопходна за развој биометријских система.

3.2.1. Историјат машинског учења

Термин “машинско учење” сковао је 1959. године Артур Семјуел, пионир у области компјутерских игара и вештачке интелигенције. У овом временском периоду коришћен је и синоним “самоучећи рачунари” (Semuel, 1988).

Најутицајније истраживање машинског учења током шестдесетих, била је Нилсонова књига “Learning Machines: Foundations of Trainable Pattern-Classifying Systems”, која се углавном бавила машинским учењем за класификацију образаца. Током седамдесетих, интересовање везано за препознавање образаца се наставило, те имамо истраживање Дуде и Харта из 1973. године “Pattern classification and scene analysis”.

1981. године, почела је имплементација коришћења наставних стратегија тако да неуронска мрежа научи да препозна 40 знакова (26 слова, 10 цифара и 4 специјална симбола) са рачунарског терминала.

Том М. Митчел је дао широко цитирану, формалнију дефиницију алгоритама који се проучавају у области машинског учења: „Компјутерски програм учи из искуства Е у односу на неку класу задатака Т и меру учинка П, уколико се његов учинак на задацима у Т, мерено са П, побољшава искуством Е" (Mitchell, 1986). Ова дефиниција задатака нуди фундаментално оперативну дефиницију уместо да дефинише машинско учење у теоретском смислу.

Ово је био следећи корак наспрам Тјуринговог предлога у раду "Рачунарске машине и интелигенција". Питање "Могу ли машине да мисле?" се замењује питањем „Могу ли машине даураде оно што ми (са способношћу мишљења) можемо да урадимо?"(Semuel, 1988).

3.2.2. Концепт и алгоритми машинског учења и приступи машинском учењу

Основни концепт машинског учења подражава коришћење статистичког учења и метода оптимизације које омогућавају рачунарима да анализирају скупове података и идентификују обрасце унутар њих. Технике машинског учења користе интелигентно тражење података тј. "Рударење података" (data mining) за идентификацију претходних начина, приликом стварања будућих модела.

Типичан надгледани алгоритам машинског учења састоји се од три компоненте (Bzdok, Krzywinski & Altman 2018):

1. Процес одлучивања
    - Рецепт или систем прорачуна који узимају податке као улазни модел, а враћају предвиђање о врсти обрасца у подацима које алгоритам жели да пронађе.
2. Функција грешке
    - Метода мерења валидности и тачности претпоставке упоређивањем са познатим примерима (када су доступни). Да ли је процес

одлучивања био исправан? Ако не, како се квантификује „колико је лош“био промашај?

3. Процес ажурирања или оптимизације
    - Алгоритам гледа на промашај, а затим се ажурира како се процес одлучивања приводи крају, тако да следећи пут промашај не буде тако велики.

Најбоље објашњење за примену ових компоненти је на примеру препоручивања филмова на некој од платформи, попут Нетфликса. Процес одлучивања алгоритма посматра колико је дати филм сличан другим филмовима које је корисник гледао и да смишља систем процењивања за различите функције.

Током процеса обуке, тј тренирања, алгоритам пролази кроз филмове које је корисник погледао и разазнаје различите особине: Да ли је то научнофантастични филм? Да ли је у питању комедија?, итд. Алгоритам затим тестира да ли заиста препоручује филмове које је корисник заправо погледао. Ако то уради како треба, функције које је користио остају исте; ако погреши, функције које су довеле до погрешне одлуке се одбијају да алгоритам не би поновио такву грешку.

Пошто се алгоритам машинског учења ажурира самостално (тако је истрениран), аналитичка тачност се побољшава са сваким следећим покретањем док се учи из података које анализира. Ова интерактивна природа учења је и јединствена јер се дешава без људске интервенције — пружајући могућност откривања скривених увида без посебног програмирања или специфичног упута од стране човека.

3.2.3. Приступи машинском учењу

Приликом машинског учења, машина учи из прошлих искустава (уноса претходних података) у односу на неку класу задатака, ако се њен учинак у датом задатку побољшава са сваким наредним искуством (IBM, 2020). На пример, претпоставимо да машина мора да предвиди да ли ће купац купити одређени производ. Машина ће то урадити гледајући претходно знање/прошла искуства,

односно податке о производима које је купац куповао сваке године и ако сваке године купује исти производ, онда постоји велика вероватноћа да ће га купац купити и ове године такође.

На концептуалном нивоу, приступи машинском учењу могу се поделити у четири типа:

- Учење под надзором
- Учење без надзора
- Учење под полу-надзором
- Учење са појачањем

3.2.4. Учење под надзором

Учење под надзором, познато и као надгледано машинско учење, користи означене скупове података за обуку алгоритама који класификују податке или тачно предвиђају резултате. Како се улазни подаци уносе у модел, он прилагођава своје функције док се модел не уклопи на одговарајући начин. Ово се дешава као део процеса унакрсне валидације како би се осигурало да модел избегне претерано или недовољно уклапање. Учење под надзором помаже организацијама да реше различите проблеме из стварног света у великом обиму, као што је класификовање нежељене поште у посебан фолдер из вашег пријемног сандучета. Неке методе које се користе у надгледаном учењу укључују: неуронске мреже (neural networks), наивни Бајес (Naive Bayes), линеарну регресију (linear regression), логистичку регресију (logistic regression), случајну шуму (random forest), машину вектора подршке (suppprt vector machine) и још много тога (IBM, 2020).

Учење под надзором може се раздвојити на два типа проблема приликом интелигентног тражења података (IBM, 2020): класификација и регресија.

- Проблеми са класификацијом користе алгоритам за прецизно додељивање тестних података у одређене категорије. Алгоритми учења под надзором могу да се користе за класификовање нежељене поште у посебан фолдер из

корисниковог пријемног сандучета. Линеарни класификатори, машине за подршку векторима, стабла одлучивања и случајна шума су уобичајени типови класификационих алгоритама.

- Регресија је још један тип метода надгледаног учења који користи алгоритам за разумевање односа између зависних и независних варијабли. Регресиони модели су корисни за предвиђање нумеричких вредности на основу различитих тачака података, као што су пројекције прихода од продаје за једно правно лице. Неки популарни алгоритми регресије су линеарна регресија, логистичка регресија и полиномска регресија.

3.2.5. Учење без надзора

Учење без надзора користи алгоритме машинског учења за анализу и груписање хетерогених скупова података. Ови алгоритми откривају скривене обрасце или груписање података без потребе за људском интервенцијом (IBM, 2020). Његова способност да открије сличности и разлике у информацијама чини га идеалним решењем за истраживачку анализу података, стратегије унакрсне продаје, сегментацију купаца, препознавање слика и образаца. Такође се користи за смањење броја карактеристика у моделу кроз процес смањења димензионалности; анализа главних компоненти и декомпозиција сингуларних вредности су два уобичајена приступа за ово. Други алгоритми који се користе у учењу без надзора укључују неуронске мреже, груписање средњих вредности, методе вероватноће кластера, итд (IBM, 2020).

Модели учења без надзора се користе за три главна задатка - груписање, асоцијацију и смањење димензионалности (IBM, 2020):

- Груписање је техника интелигентног тражења података за груписање неозначених података на основу њихових сличности или разлика. На пример, K-means алгоритми груписања додељују сличне скупове података групама, где K вредност представља величину груписања и грануларност. Ова техника је корисна за сегментацију тржишта, компресију слике, итд.

- Асоцијација је још један тип метода учења без надзора који користи различита правила за проналажење односа између варијабли у датом скупу података. Ове методе се често користе за анализу потрошача и модела за интернет препоруке, у складу са принципом - „Купци који су купили артикал А, такође су купили артикал Б“.
- Смањење димензионалности је техника учења која се користи када је број карактеристика (или димензија) у датом скупу података превисок. Смањује број уноса података на величину којом се може управљати уз истовремено очување интегритета података. Често се ова техника користи у фази пред процесирање података, као што је када аутоматски кодери уклањају шум из визуелних података да би побољшали квалитет слике.

3.2.6. Учење под полу-надзором

Основна разлика између учења под надзором и учења без надзора је у томе што скупови података за учење под надзором имају излазну ознаку повезану са сваком структуром података која се састоји из више јединица (tuple), док скупови података за учење без надзора немају.

Најосновнији недостатак било ког алгоритма за надгледано учење је то што скуп података мора да буде мануелно хомогенизован или од стране инжењера машинског учења или стручњака за податке (Introduction to Semi-Supervised Learning, 2022). Ово је веома скуп процес, посебно када се ради о великим количинама података. Најосновнији недостатак било ког ненадгледаног учења је тај што је његов спектар примене ограничен.

Да би се супротставили овим недостацима, уведен је концепт полунадгледаног учења. У овој врсти учења, алгоритам се обучава на комбинацији хомогених и хетерогених података. Обично ће ова комбинација садржати веома малу количину хомогених података и веома велику количину хетерогених података. Основна процедура која је укључена је да прво програмер групише сличне податке користећи алгоритам учења без надзора, а затим користи постојеће

хомогене податке да означи остатак хетерогених података (Introduction to Semi-Supervised Learning, 2022). Типични случајеви употребе оваквог типа алгоритма имају заједничку особину међу њима – аквизиција хетерогених података је релативно јефтина, док је хомогенизовање наведених података веома скупо (Introduction to Semi-Supervised Learning, 2022).

Замислимо три типа алгоритама учења као учење под надзором где је ученик под надзором наставника и код куће и у школи, учење без надзора где ученик мора сам да смисли концепт и полунадгледано учење где наставник предаје неколико концепата на часу и даје питања као домаћи задатак која су заснована на сличним концептима.

Полу-надзирани алгоритам претпоставља следеће о подацима (Introduction to Semi-Supervised Learning, 2022):

- Претпоставка континуитета: Алгоритам претпоставља да је већа вероватноћа да тачке које су ближе једна другој имају исту излазну ознаку.
- Претпоставка кластера: Подаци се могу поделити у дискретне кластере и већа је вероватноћа да ће тачке у истом кластеру делити излазну ознаку.
- Претпоставка многострукости: Подаци леже приближно на многострукости много ниже димензије од улазног простора. Ова претпоставка дозвољава коришћење растојања и густина које су дефинисане на многострукости.

3.2.7. Учење са појачањем

Учење са појачањем је много разноврсније од приступа учењу под надзором и без надзора. Учење методично учи да изврши задатак и тренира алгоритме користећи систем награђивања и казне.

Постоје софтверски агенти који добијају "награде" након прецизног обављања функције и "казне" за нетачан учинак. Агент је програмиран на такав начин да не захтева никакву људску интервенцију. Учење са појачањем се фокусира на акцију праћену побољшаним перформансама, или се радња поново не спроводи у

пракси. За сада постоји неколико ограничења за учење са појачањем, која се могу решити кроз учење са дубоким појачањем (Bajaj, 2021).

Постоје две врсте појачања (Bajaj, 2021):

1. Позитивно

Позитивно појачање дефинише да се догађај, који се догоди због одређеног понашања, повећава снагу и учесталост понашања. Другим речима, позитивно утиче на понашање.

Предности учења са појачањем су:

- Извлачи максимум из системских перформанси;
- Подржава промене система током дужег временског периода;
- Превише појачања може довести до преоптерећења стања што може умањити резултате.

2. Негативно

Негативно појачање се дефинише као јачање понашања јер се негативно стање зауставља или избегава.

Предности учења са појачањем:

- Повећава перформансе понашања;
- Пркоси минималном стандарду учинка;
- Обезбеђује само довољно да испуни минимално понашање.

Према студијама, примећено је да је синтеза алгоритама машинског учења са биометријом обезбедила бољу биометријску тачност у поређењу са конвенционалним методама.

4. Примена машинског учења и биометрије понашања на безбедност веб апликација

4.1. Аутентификација као тип безбедности у веб свету

Пошто се аутентификација одвија тренутно и обично само једном, превара идентитета је могућа. Нападач може заобићи биометријски систем аутентификације. Украден биометријски систем представља тежак проблем. За разлику од лозинки или паметних картица, које се могу променити или поново издати, отисак прста или изглед ока су трајне карактеристике. Када нападач успешно ископира те карактеристике, крајњи корисник је у потпуности искључен из система, што повећава безбедносне ризике и/или трошкове поновне имплементације. Статичке физичке карактеристике могу се дигитално копирати, на пример, лице се може копирати помоћу фотографије, отисак гласа помоћу снимка гласа, а отисак прста помоћу различитих метода (лепљење, скулптурисање, итд.) (Bonaccorso, 2017). Поред тога, статичка биометрија је понекад нетолерантна на промене у физиологији као што су дневне промене гласа или промене изгледа (шминкање, на пример) (Bonaccorso, 2017).

Унимодални биометријски системи (системи који користе само један тип биометрије) морају да се суоче са разним проблемима као што су "бучни" подаци (loud data), варијације унутар класе, ограничени степени слободе приступа, неуниверзалност, лажни напади и неприхватљиве стопе количине грешака (Bonaccorso, 2017). Нека од ових ограничења могу се решити применом мултимодалних биометријских система (системи који користе више типова биометрије симултано) који интегришу доказе представљене из више извора информација. Мултимодални биометријски системи су поузданији због присуства вишеструких, независних података. Ови системи су у стању да задовоље строге захтеве перформанси које намећу различите апликације. Они се баве проблемом неуниверзалности, пошто рударење података покрива велику већину становништва, па је количина прикупљених податак огромна. Они такође

спречавају лажирање јер би нападачу било тешко да ископира више биометријских особина правог корисника симултано (Bonaccorso, 2017). Штавише, они могу осигурати да је иницијални, прави корисник заиста присутан на месту аквизиције података, тиме што могу захтевати презентацију биометријских особина насумично.

У наредном поглављу су изведени принципи који се генерално користе кроз принципе машинског учења у дирекној корелацији са биометријом понашања.

4.2. Коришћење биометрије понашања кроз принципе машинског учења

Већ знамо да је „биометријска технологија“систем који верификује идентитет појединца у реалном времену, користећи његове јединствене карактеристике као што су отисци прстију и геометрија лица. Сматра се да је интеграција машинског учења са биометријском технологијом револуционирала свет технологије.

Машинско учење користи огромну количину структурираних података тако да може да генерише резултате или даје предвиђања на основу датих података. Тренутно се машинско учење користи у алгоритмима Гугл претраге, филтерима за нежељену пошту, предлагања Фејсбук пријатеља и препорука за интернет куповину (Bonaccorso, 2017).

Током банковног трансфера, биометрија понашања може анализирати притиске на тастере урачунавајући брзину куцања и редослед прстију који се користе за унос, а у року од 10 минута може да креира профил који је довољно јак да потврди корисника.

Међутим, како време пролази, а особа све чешће користи уређај, њено понашање се мења и прилагођава. Машинско учење помаже у пробијању мешања и збуњивања различитих сигнала и проналази конзистентност у обрасцима понашања током времена, без обзира на промене.

У данашње време, скоро 100% превара долази од саме аутентификације и дешава се када се легитимни корисник пријави, али рачун преузима малвер вирус,

ботови, коришћење социјалног инжењеринга и други типови напада на даљину (Bonaccorso, 2017).

Приступи машинског учења помажу у идентификацији, функцијама препознавања и систематизацији података која је неопходна за развој биометријских система. Као што је већ напоменуто, приступи машинском учењу могу се поделити у четири типа: учење под надзором, учење без надзора, учење под полу-надзором, учење са појачањем. Сваки од ових типова има другачију интеракцију са биометријом понашања и може се користити у различите сврхе. Пошто је учење под полу-надзором релативно нова дисциплина, традиционални приступ интергације биометрије понашања у машинско учење се базира на учењу под надзором, учењу без надзора и учењу са појачањем.

4.2.1. Биометрија и учење под надзором

Учење под надзором је помогло неколико биометријских апликација кроз велики број алгоритама. Неки од алгоритама су конволуцијске неутралне мреже, методе кернела, логистичка регресија и стабла одлучивања (Supervised and Unsupervised Learning, 2022).

Примена:

- Препознавање лица
- Класификација емоција у говору
- Препознавање емоција лица

4.2.2. Биометрија и учење без надзора

Ненадгледано машинско учење имплементирано у биометрији обезбеђује боље методе учења, омогућава бољу класификацију и тачно позиционирање биометријских карактеристика. Учење без надзора се може користити за извлачење потпуног аутоматског узорка вена прста. Међутим, може се користити само у прелиминарној фази за боље дефинисање стратегије учења, синтезу

карактеристика, анализу података итд (Supervised and Unsupervised Learning, 2022).

Примена:

- Учење без надзора у узорку вена прста - Потпуно аутоматско вађење узорка вена прста.
- Учење без надзора у препознавању отиска прста
- Упаривање ретиналног узорка – Сегментација крвних судова мрежњаче.
- Детекција гласа

4.2.3. Биометрија и учење са појачањем

Трећи приступ, тј. учење са појачањем, много је разноврснији од приступа учењу под надзором и без надзора. Софтверски агент аутоматски прима награде за исправан рад и казне за нетачан рад. Агент је програмиран да одређује без интервенције човека само максимизирањем своје компензације и минимизирањем своје казне (Bajaj, 2021).

Према студијама, примећено је да је фузија алгоритама машинског учења са биометријом обезбедила бољу биометријску тачност у поређењу са конвенционалним методама.

Примена:

За разлику од прва два приступа, учење са појачањем је најпримењивије уз биометрију и није ограничено на одређени сет података. Примена је широка - од аутоматизације у свим индустријама, нарочито аутомобилској, трговини и финансијама, НЛП-у, медицини, препоручивању садржаја, видео играма, маркетингу, итд (Bajaj, 2021).

4.3. Значај безбедности приступа веб апликацијама

У ери у којој интернет трансакције подстичу продају и генеришу огроман приход за и интернет предузећа и општа правна лица, неовлашћени и злонамерни хакерски напади и упади у систем, коштају компаније милионе долара.

Како вредност и употреба информација наставља да расте, појединци и предузећа траже додатне начине за обраду и складиштење података. Најчешће се користе системи за интелигентну обраду података (Khan, Sajin & Chakravarthy, 2018).

Систем за руковање подацима генерално обрађује, сакупља, складишти и/или преводи чисте информације или податке у пословне и личне податке (усмерене или фокусиране) и на тај начин омогућава корисницима да искористе вредност података.

Варијације у системима за руковање подацима омогућавају да системи за руковање подацима буду општи или конфигурисани за одређеног корисника или специфичну употребу, као што су обрада финансијских трансакција, резервације авио-компанија, складиштење података предузећа или глобалне комуникације (Khan, Sajin & Chakravarthy, 2018). Поред тога, системи за руковање информацијама могу укључивати различите хардверске и софтверске компоненте које се могу конфигурисати за обраду, складиштење и преношење информација и могу укључивати један или више рачунарских система, система за складиштење података и мрежних система.

Савремени системи за руковање информацијама обухватају много различитих типова потрошачких и комерцијалних електронских уређаја као што су, на пример, лични рачунари (нпр. стони или лаптопови), таблет рачунари, мобилни уређаји (паметни телефони) корпоративне (или предузетничке) сервере и системе за обраду података и слично. Ови уређаји могу да се разликују по величини, облику, перформансама, функционалности и цени. У сваком случају, скоро сви ови модерни уређаји су опремљени релевантним хардвером и софтвером који омогућавају њиховим корисницима приступ великом броју различитих интернет локација, као и обављање интернет трансакција.

У свету којим доминирају е-трговина и електронске трансакције, пословна вредност безбедне веб странице је новчано немерљива. Са широко распрострањеном доступношћу вештачке интелигенције и прикупљања података

путем мрежних ботова, хакери имају на располагању далеко софистицираније алате за лажно пријављивање на веб-сајтове, чиме се организују крађе идентитета на порталима за пријаву (Khan, Sajin & Chakravarthy, 2018). Ово чини корисничке податке подложним злоупотреби и, када су акредитиви за пријаву корисника компромитовани, уљези могу да изврше вишеструка наредна злонамерна пријављивања која остају практично неоткривена током аутентификације за пријављивање (Khan, Sajin & Chakravarthy, 2018).

4.4. Безбедносни модел машинског учења и биометрије понашања

Безбедносни модел машинског учења и биометрије понашања се односи на заштиту од крађе идентитета у онлајн трансакцијама, тј. на безбедносни модел веб локације у којем је биометрија понашања корисника – као што су динамика померања курсора, притискање тастера и шаблон кликова – током пријављивања на веб локацију, прикупљена за креирање и имплементацију прилагођеног безбедносног модела заснованог на вишеструким класификаторима машинског учења. Поента иза имплементације овог модела је како би се корисник разликовао од имитатора и како би се обезбедила унапређена безбедносна структура за пријављивање на веб локацију корисника чак и ако су акредитиви за пријаву корисника угрожени.

4.4.1. Типови биометријских података за аутентификацију на веб апликацијама

Истраживања о безбедносним апликацијама које користе информације о понашању клијента за аутентификацију су идентификовала два потенцијална извора релевантних података:

1. покрете миша корисника и
2. притиске тастера.

Ове апликације користе евиденције покрета миша и притисак на тастере у изолацији или у елементарној корелацији, као и у ограниченом капацитету. На пример, неке апликације користе такву биометрију понашања за поновну

аутентификацију — као што је захтевање поновног уноса лозинке — али не као први зид или корак безбедности. Неке друге апликације користе двоструки сигурносни систем да замене конвенционалне начине пријаве, попут лозинке.

У овим апликацијама, шаблон заснован на притискању тастера је први ниво аутентификације, а покрети миша су други. Међутим, такав приступ се не заснива на пасивној аутентификацији, где биометрија функционише како би допунила постојеће безбедносне протоколе. Уместо тога, овај двоструки безбедносни приступ користи клијентов шаблон за притискање тастера као ентитет за који се обезбеђује аутентификација (Khan, Sajin & Chakravarthy, 2018). Слично, шаблон јединственог покрета миша се користи као „лозинка“за улаз на други ниво безбедности. Наравно, неке друге апликације користе биометријски модел да додају додатне слојеве безбедности—на пример, за превенцију губитка података предвиђањем идентитета креатора података или за идентификацију корисничког профила у веб апликацијама (Khan, Sajin & Chakravarthy, 2018).

Стога, као што је горе наведено, тренутне безбедносне апликације користе значајно ограничен скуп биометријских података корисника — наиме, покрете миша и притиске тастера. Као резултат тога, карактеристике извучене из биометрије понашања клијента су такође ограничене. На пример, већина апликација које анализирају покрете миша идентификују класе (најчешће 8 класа) у које се сваки догађај миша може доделити, на основу релативног смера кретања миша (Khan, Sajin & Chakravarthy, 2018).

Са друге стране, биометрија притиска на тастер се углавном фокусира на време задржавања (време када тастер остаје притиснут) и време преласка (време између притискања „тастер на горе“и следећег „тастера на доле“).

Штавише, прикупљање биометријских података може се проширити изван активности пријављивања, на пример, уз помоћ софтверског агента који се налази на рачунару корисника. Такво проширено прикупљање података може

резултирати великом количином података које треба анализирати, што је често захтевно и скупо за компаније.

4.4.2. Типови алгоритама за аутентификацију на веб апликацијама

Тип алгоритама за обраду биометријских података које користе тренутне безбедносне апликације разликује се од апликације до апликације (Attaie, Caldwell, Ward, Yassin, Graham, & Elliot, 2019).

- Једна врста апликација креира класификатор који верификује сличност између шаблона који треба да се верификује и шаблона прототипова (направљених од прикупљених евиденција биометријских података корисника) користећи мере сличности засноване на удаљености између вектора карактеристика-образац-прототип.
- Друге врсте апликација користе посебну подршку класификатора векторске машине за динамичке карактеристике миша и за притисак на тастере.
- Остале апликације користе једну од следећих метода за класификацију: Бајесове мреже, ауто-асоцијативна неуронска мрежа, Монте Карло приступ, Гаусова функција густине вероватноће, мера сличности правца или паралелна стабла одлучивања.

Примећено је да су се перформансе ових алгоритама разликовале на основу њихових инхерентних предрасуда и варијабли (у зависности од претходног учења и типа података које су обрађивале). Као резултат тога, перформансе детектора аномалија (за идентификацију пријављивања имитатора) можда неће бити доследне у свим таквим апликацијама или међу више корисника у оквиру исте апликације (Attaie, Caldwell, Ward, Yassin, Graham, & Elliot, 2019).

4.4.3. Биометрија понашања корисника за обезбеђивање пријављивања на веб локацију

Биометрија понашања корисника током пријављивања на веб локацију укључује клијентову динамику кретања курсора, притиске тастера и обрасце кликова мишем (Attaie, Caldwell, Ward, Yassin, Graham, & Elliot, 2019).

У једној варијанти модела, ове биометријске карактеристике се користе за креирање прилагођеног безбедносног модела заснованог на машинксом учењу за сваког корисника. Овај прилагођени безбедноснни модела може да разликује инциијалног корисника од имитатора (било да је у питању човек или истренирани систем) (Khan, Sajin & Chakravarthy, 2018). Такав модел, у комбинацији са постојећим безбедносним протоколима, може да обезбеди побољшану безбедност за профил корисника током пријављивања, чак и ако су акредитиви за пријаву корисника компромитовани или хаковани. Модел може у почетку да прикупља релевантне биометријске податке о понашању са стране корисника када корисник креира нови налог на веб локацији или када се корисник првобитно пријави на веб локацију (Khan, Sajin & Chakravarthy, 2018).

Програмски код или софтверски модул за прикупљање података може се лако интегрисати са било којом веб апликацијом без утицаја на перформансе веб локације. Прикупљени биометријски подаци могу се користити за обуку колекције класификатора заснованих на машинксом учењу у безбедносном моделу.

Колекција класификатора може укључивати следећа три класификатора (Khan, Sajin & Chakravarthy, 2018):

1. Класификатор заснован на вишеслојном перцептрону (вештачка неуронска мрежа која генерише скуп излаза из скупа улаза. Карактерише неколико слојева улазних чворова повезаних као усмерени индекс између улазног и излазног слоја),
2. Класификатор заснован на машини векторске подршке и
3. Класификатор заснован на адаптивном појачавању (AdaBoost).

Након довољног броја прикупљених података за пријаву, модел може да примени средства за откривање биометријски базираних превара како би заштитио налог корисника од будућих имитатора, тј. хакера. Откривање преваре се може одрадити преко обучених верзија класификатора, који се упоређују да би се дало оптимално предвиђање у реалном времену (док се корисник пријављује). (Khan, Sajin & Chakravarthy, 2018)

4.4.4. Метода машинског учења за обезбеђивање пријављивања на веб локацију

Метода машинског учења за обезбеђивање пријављивања на веб локацију обухвата три корака (Khan, Sajin & Chakravarthy, 2018):

1. Пријем скупа података (од стране рачунарског система) специфичног за првог корисника, када се први корисник пријави на веб локацију, при чему скуп података укључује следеће:
    - Координате курсора у сваком тренутку када први корисник помери курсор током пријављивања,
    - временске ознаке повезане са притиском и отпуштањем сваког тастера на уређају за унос података од стране првог корисника током пријављивања,
    - временске ознаке свих кликова мишем од стране првог корисника (од почетка до краја пријављивања), и
    - специфичан код за сваки тастер који је први корисник притиснуо на тастатури, током пријављивања.

2. Тренирање више класификатора (помоћу рачунарског система) заснованих на машинском учењу, у моделу машинског учења заснованом на примљеном скупу података специфичном за корисника;

3. Коришћење (од стране рачунарског система) обученог модела машинског учења за одбијање покушаја пријављивања на веб локацију од стране другог корисника са акредитивима за пријаву првог корисника, при чему се други корисник разликује од првог корисника. У посебним ситуацијама, мноштво

класификатора заснованих на машинском учењу је комбинација горе наведених класификатора.

Пошто су алати за хаковање засновани на вештачкој интелигенцији све доступнији за крађе података, компаније се скоро свакодневно суочавају са изазовима попут напада на њихове веб локације и корисничке базе. Без интегрисаног безбедносног система, запослени на обезбеђењу веб-сајта све теже одолевају нападима хакера, јер не могу да интервенишу у реалном времену када су акредитиви за пријаву корисника компромитовани. Због тога, безбедносни модел који користи машинско учење базорано на интеграцији биометријских података има веће шансе да обезбеди виши степен безбедности компанијама (Attaie, Caldwell, Ward, Yassin, Graham, & Elliot, 2019).

Обука комбинованих класификатора заснованих на машинском учењу, са прикупљеним биометријским подацима о понашању и накнадним испитивањем комбинованих класификатора обезбеђује оптимално предвиђање у реалном времену са релативно малом количином података, обезбеђујући да безбедносни модел који је јефтин и ефикасан. Као резултат тога, откривање превара/крађе у реалном времену може да се постигне без утицаја на перформансе пријављивања на одговарајућу веб локацију (Khan, Sajin & Chakravarthy, 2018).

4.4.5. Перформансе изабраних класификатора у односу на скуп метрика учинка

Тачност класификатора се може дефинисати као пропорција правилно класификованих узорака у укупном броју узорака. Пошто се тачност односи на збир перформанси класификатора, модел машинског учења има различито понашање на свакој класи. Када се обезбеђује пријављивање на веб локацију, учесталост узорака међу класама корисника можда неће бити уједначена - може бити много позитивних узорака (правих пријава корисника) и врло мало негативних узорака (лажни корисник који покушава да се пријави) (Khan, Sajin & Chakravarthy, 2018).

Ако се процењује само тачност класификатора, понекад није могуће тачно одредити колико исправно модел идентификује лажне пријаве јер ће перформансе на истинитим пријавама имати много већи допринос тачности и тачност ће бити блиска учинку на позитивној класи.

Стога, тачност се не узима као једина метрика учинка класификатора, него се посматра заједно са прецизношћу и опозивом (recall). На пример, у неким случајевима, перформансе класификатора могу бити прецизне, али не и тачне. Процена перформанси класификатора у односу на више метрика перформанси може да обезбеди свеобухватнији модел машинског учења за откривање превара (Khan, Sajin & Chakravarthy, 2018).

Прецизност као метрика перформанси се карактерише (за једну, одређену класу) као однос правих пријава за дату класу према броју укупно предвиђених пријављених корсника.

Опозив као метрика перформанси се карактерише (за једну, одређену класу) као однос правих пријава за класу наспрам укупног броја стварних пријава за дату класу (Khan, Sajin & Chakravarthy, 2018).

Обе ове метрике, а нарочито у комбинацији, дају много бољи увид у то да ли модел заправо тачно идентификује лажне пријаве или не.

5. Закључак

Дигитална ера започета комерцијализацијом рачунара за кућну употребу условила је људско друштво на промењен начин прикупљања, обраде и складиштења података.

Како нас само један клик дели од информација попут оних о банковним рачунима, медицинским и универзитетским документима, државним списима и документима и слично, приступ истим се мора верификовати, а индентификација корисника мора бити стабилна и континуирана.

Машинско учење је способно да учи из људског понашања на основу скупа података биометрије понашања и континуирано побољшава корисничке профиле који се могу користити за аутентификацију приступа или трансакција. На пример, током банковног трансфера, биометрија понашања може анализирати притиске на тастере гледајући брзину куцања и који прсти се користе за куцање, а у року од 10 минута може да креира профил који је довољно јак да потврди идентитет корисника. Међутим, како време пролази, а особа све чешће користи уређај, њено понашање се мења и прилагођава. Машинско учење помаже у раздвајању различитих сигнала и лоцира доследност у обрасцима понашања током времена, без обзира на промене или спољашње факторе.

Иако је примена машинског учења (и вештачке интелигенције) у ове сврхе почела ради обезбеђивања финансијских трансакција, све више је јасно да ће се примена комбинације машинског учења и биометрије понашања евалуирати и распространити на многе индустрије и научне гране.

Што се тиче предвиђања за еволуцију овог модела у сајбер безбедности, три најчешће помињане предикције од стране истраживачких кућа из домена сајбер безбедности су:

1. Да ће степен коришћења ових технологија наставити да расте и да се примењује;

2. Да ће почети да се примењује аутоматизовано откривање аномалија (не само приликом пријаве на веб локацију него и током коришћења);
3. Да ће сајбер безбедност прећи из реактивног у проактивно стање.

Машинско учење отвара до сада незамисливе могућности за аутоматизацију, ефикасност и иновације. Са развојем технологије за надзор, биометријски подаци су не само доступни него и уснимљени и лако мерљиви. Модалитет ове комбинације је флексибилан, ефикасан, безбедан и јефтин за примену.

6. Литература

Attaie, T., Caldwell, J., Ward, T., Yassin, Y., Graham, J., & Elliot, K. (2019, October). *Dynamic keystroke for authentication with machine learning algorithms.*

Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2018). *Challenges and applications in multimodal machine learning.* In The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition-Volume 2

Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.

Bzdok, D., Krzywinski, M., & Altman, N. (2018*). Machine learning: supervised methods*. Nature methods

Cao, H. (2020*). Remote Gait Monitoring Mobile System Enabled by Wearable Sensor Technology* (Doctoral dissertation, Case Western Reserve University)

De Souza, J.W.M., Medeiros, A.G., Holanda, G.B., Rego, P.A.L., Rebouças Filho, P.P. (2022). *Fingerprint Classification Based on the Henry System via ResNet*. In: Rozinaj, G., Vargic, R. (eds) Systems, Signals and Image Processing. IWSSIP 2021. Communications in Computer and Information Science, vol 1527. Springer, Cham. https://doi.org/10.1007/978-3-030-96878-6_2

Falaah Arif Khan, Sajin Kunhambu, K. Chakravarthy G, 6th International Symposium, (2018), *Behavioral Biometrics and Machine learning to Secure Website Logins*. Bangalore, India, September 19–22, 2018, Revised Selected Papers

GeeksForGeeks (2019, April 15), *Introduction to Deep Learning*. Geeksforgeeks website.

GeeksForGeeks (2021, November 18), *Supervised and Unsupervised Learning*. Geeksforgeeks website. Supervised and Unsupervised learning - GeeksforGeeks

Hitachi LTD, (2005, March 31), *Annual Report 2005*.Hitachi LTD website. ANNUAL REPORT 2005 (hitachi.com)

IBM, (2020, August 19), *Machine Learning, Supervised Learning*. IBM Cloud Education. What is Supervised Learning? | IBM

IBM, (2020, September 21), *Machine Learning, Unsupervised Learning*. IBM Cloud Education. What is Unsupervised Learning? | IBM

Introduction to Semi-Supervised Learning,(2022, March, 29) Introduction to Semi-Supervised Learning - Masterful 0.4.1 documentation (masterfulai.com)

Jordan, M. I., & Mitchell, T. M. (2015). *Machine learning: Trends, perspectives, and prospects*.

Kochegurova, E. A., Gorokhova, E. S., & Mozgaleva, A. I. (2017). *Development of the keystroke dynamics recognition system*. In Journal of Physics: Conference Series (Vol. 803, No. 1, p. 012073). IOP Publishing.

Markowitz, J. A. (2000). *Voice biometrics*. Communications of the ACM

Mayhew S., (2018, February 1). *History of Biometrics*. Biometrics update website. https://www.biometricupdate.com/201802/history-of-biometrics-2

McCarthy, J. (2007, November 12). *What is artificial intelligence?* Stanford University, whatisai.pdf (stanford.edu)

Monrose, F., & Rubin, A. D. (2000). *Keystroke dynamics as a biometric for authentication*. Future Generation computer systems, 16 (4)

Prateek Bajaj, (2021, November 18), *Reinforcement Learning*. Geeksforgeeks. Reinforcement learning - GeeksforGeeks

Russell, S., & Norvig, P. (2002). *Artificial intelligence: a modern approach*. The fourth edition, 2020.

Salekar K. (2022) *How Artificial Super Intelligence Works and applies.* Technohealth. How Artificial Super Intelligence Works & Applied | Techohealth.com

Samuel, A. L. (1988). *Some studies in machine learning using the game of checkers*. II—recent progress. Computer Games I

Simeón, F., & Monge, J. C. (2005). Kinesiology. Revista de Enfermeria (Barcelona, Spain), 28(12),

Trustwave (2018, April 05). *2018 Trustwave Global Security Report*. Trustwave website. 2018 Trustwave Global Security Report | Trustwave

Wang, L., & Geng, X. (Eds.). (2009). *Behavioral Biometrics for Human Identification: Intelligent Applications.* Intelligent Applications. IGI Global.

A10 Networks company research team, (2016, December 19). *10 with A10: 10 Cyber Security Predictions for 2017*. A10networks website. https://www.a10networks.com/blog/cyber-security-predictions-2017/

И З Ј А В А О А К А Д Е М С К О Ј Ч Е С Т И Т О С Т И

Изјављујем да сам у приложеном раду поштовао/ла сва правила о академској честитости.

Овај писани рад резултат је искључиво мог личног рада, темељи се на мојим истраживањима и ослања се на наведену литературу.

У Београду, дана _____ године.

Потпис студента:

_____