RIINVEST
Data Science And Business Analytics Department

Punim Seminarik

Engjëll Gashi

Inteligjenca Artificiale dhe parashikimi i shkeljes se te dhenave.

Profesori: Prof.Valon Kastrati

Tetor / 2022

Tabela e permbajtjes

1.Hyrje 3
2.Qka eshte vleresimi i rrezikut te shkeljes se te dhenave? 3
3.Qka eshte Inteligjenca Artificiale? 5
4.Inteligjenca Artificiale kunder analizes se te dhenave 6
5. Nderthurja e Inteligjences Artificiale me parashikimin e shkeljes se te dhenave. 6
6.Konkluzioni 12

1.Hyrje

Sulmet kibernetike ne bote vetem shtohen cdo dite e me shume. Varesisht nga madhesia e biznesit, ka nje numer shume te madh te sulmeve te cilat duhet te analizohen ne kohe reale te evaluohen per masen e rrezikut te cilen ato paraqesin. Edhe nese entiteti i caktuar e merr persiper te rrise vetedijen per siguri kibernetike apo te rrise masat e sigurise eshte e pamundur qe te perballen me numrin e madh te sulmeve te cilat nuk mund te trajtohen nga individe. Ka shume sulme kibernetike te cilat nuk mund te analizohen edhe nga nje shume e madhe e njerezve per shkaqe te natyres, numrit te madh dhe madhesise se sulmit. Pasi qe nje sfide e tille u eshte shfaqur ketyre entiteteve, ka qene mese e arsyeshme qe te zhvillohen metoda te reja per trajtimin e ketyre sulmeve. E kjo ka mundur te arritet me vegla te inteligjences artificiale qe bazohen ne siguri kibernetike ne menyre qe te ulen rreziqet e shkeljes se te dhenave dhe filtrimi i sulmeve “potenciale”. IA dhe “Machine Learning” (perkthim. Mesimi i makines”) ML jane bere vegla shume te rendesishme ne sigurine kibernetike, per shkak te fuqise se filtrimit, analizes dhe parashikimit te sjelljeve te ndryshme te kercenimeve, duke i analizuar dhe duke mesuar se si funksionojne ato. Kjo metodologji ka mundesuar qe keto vegla te kuptojne sjelljen e sulmeve dhe te detektojne ato ne kohe te shpejte varesisht numrit te madh te sulmeve.

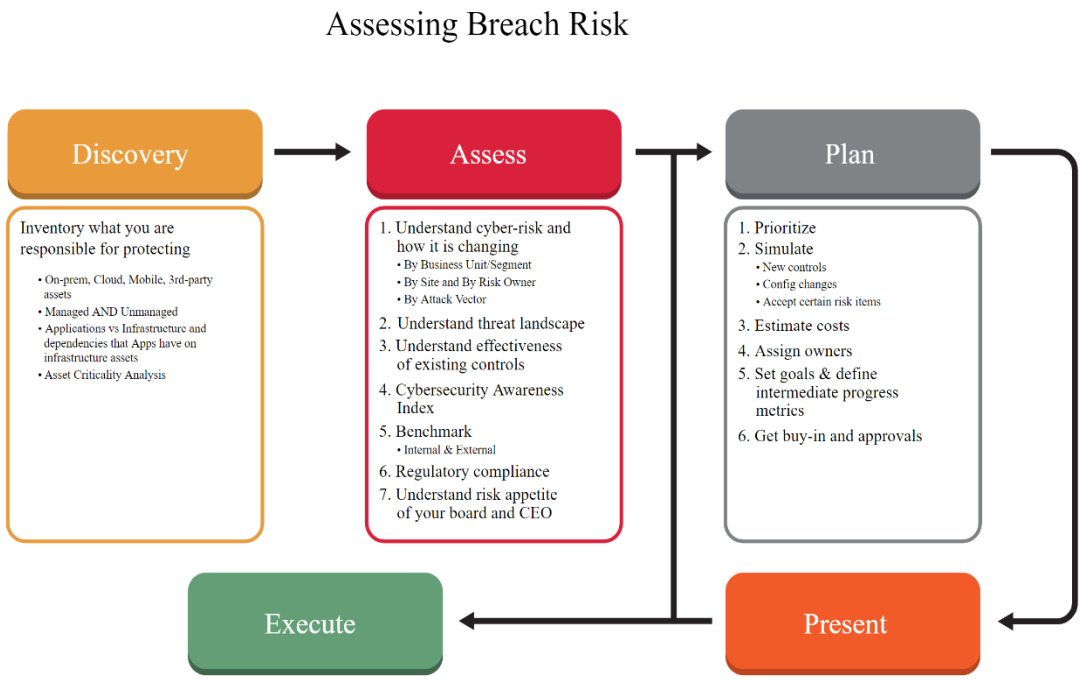
2.Qka eshte vleresimi i rrezikut te shkeljes se te dhenave?

Nje vleresim i rrezikut eshte nje proces zyrtar qe perdoret nga organizatat per te identifikuar kercenimet dhe dobesite qe mund te ndikojne negativisht ne sigurine kibernetike.[1]

Nga ana tjeter, Vleresimi i Rrezikut te Shkeljes se te dhenave eshte nje vleresim proaktiv i rrezikut, perkundrazi nje aktivitet i vetekontrolles duke marre parasysh sjelljet e shkeljeve te ndodhura ne industri te ngjashme ne te kaluaren.[2]

[1] SISA Editor, (2021), Breach Risk Assessment at *https://www.sisainfosec.com/services/breach-risk-assessment/*
[2] SISA Editor, (2021), Breach Risk Assessment at *https://www.sisainfosec.com/services/breach-risk-assessment/*

Sa i perket vleresimit te rrezikut te shkeljes se te dhenave me poshte do paraqesim nje diagram nga Balbix.



[3]

Per te bere nje vleresim te rrezikut te shkeljes duhet te ndjekim 5 hapa:

Zbulimin – Te dihen asetet te cilat jane per detyre te mbrohen.

Vleresimin – Te dime se si te vleresojme rrezikun.

Planifikimin – Te kuptohet se si do shkoje plani i vleresimit te rrezikut, kostot dhe simulime sigurie.

Prezantimin – Te prezantohen tri pikat e para tek personat pergjegjes.

Ekzekutimin – Plani i krijuar pas aprovimit te ekzekutohet.

Nese te gjitha keto pika plotesohen me kujdes, saktesi dhe efikasitet, atehere vleresimi dhe raportimi i rrezikut te shkeljes do jete shume me i lehte, me i kuptueshem dhe nuk do jete problem i madh ne te ardhmen.

---

[3] Rich Campagna, (2020, July), Making Infosec Jobs Easier: Assessing and Reporting Breach Risk, at *https://www.balbix.com/blog/infosec-jobs-assessing-reporting-breach-risk/*

3.Qka eshte Inteligjenca Artificiale?

Inteligjenca Artificiale eshte nje metode per te bere nje kompjuter ose nje produkt per te menduar ne menyre te zgjuar mendon njeriu.[4] Inteligjenca Artificiale mundohet te arrije te gjeje se si truri i njeriut funksionon, kur perpiqet te zgjidhe problemet. Keto studime nxjerrin sisteme inteligjente te softuerit.[5]

Inteligjenca eshte e paprekshme. Perbehet nga Arsyetimi, Te mesuarit, Zgjidhja e Problemeve, Perceptimi, Inteligjenca gjuhesore. Shkenca kompjuterike terheq AI ne fushen e shkences, matematikes, psikologjise, gjuhesise, filozofise etj.[6] Disa nga fushat ku AI aplikohet jane: Videolojerat, Sistemet e vizionit, kuptimi i fjaleve, kuptimi i shkrimit, robotet inteligjente.[7] Disa nga qellimet afatgjata te AI jane: Arsyetimi i dijes, planifikimi, machine learning, procesimi i gjuhes natyrale, vizioni kompjuterik, robotika.[8]

Si nje term me definitiv sa i perket inteligjences artificiale eshte se Inteligjenca artificiale (AI) i referohet simulimit te inteligjences njerezore ne makinat qe jane programuar te mendojne si njerezit dhe imitojne veprimet e tyre. Termi mund te zbatohet gjithashtu per çdo makine qe shfaq tipare te lidhura me nje mendje njerezore si mesimi dhe zgjidhja e problemeve.[9] Njeri nder shembujt me te perditshem qe mund te marrim per perdorimin e AI jane makinat qe vozisin veten, ku me llogaritje te shumta makinat mund te kuptojne rrezikun dhe te veprojne ne menyre qe ta evitojne. Makinat marrin inputin nga jashte dhe pas llogaritjeve te shumta vijne tek rezultatet te cilat nuk lejojne makinen te kete perplasje. AI perdoret edhe ne sektorin e financave, ku perdoret per te zbuluar dhe ta beje me dije (flag) aktivitetin ne banke dhe financa siç jane perdorimi i pazakonte i kartes se debitit dhe depozitat e medha te llogarise - te gjitha keto ndihmojne departamentin e mashtrimit te nje banke.[10]

---

[4] Meruja Selvamanikkam, (2018, August), Introduction to Artificial Intelligence, at *https://becominghuman.ai/introduction-to-artificial-intelligence-5fba0148ec99*

[5] Meruja Selvamanikkam, (2018, August), Introduction to Artificial Intelligence, at *https://becominghuman.ai/introduction-to-artificial-intelligence-5fba0148ec99*

[6] Meruja Selvamanikkam, (2018, August), Introduction to Artificial Intelligence, at *https://becominghuman.ai/introduction-to-artificial-intelligence-5fba0148ec99*

[7] Lbid

[8] Lbid

[9] Jake Frankenfield, (2021, March), Artificial Intelligence (AI), at *https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp*

[10] Jake Frankenfield, (2021, March), Artificial Intelligence (AI), at *https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp*

4.Inteligjenca Artificiale kunder analizes se te dhenave

AI eshte nje fjale shume e njohur koheve te fundit e cila edhe keqperdoret. Si edhe Big Data, cloud, IoT etj, nje shume ndermarrje po kerkojne menyra per t'u hedhur tek fusha e AI. Edhe pse perdorin teknologji qe analizojne te dhenat dhe lejojne qe rezultatet te vendosin rezultate te caktuara, kjo nuk eshte AI; AI i paster ka te beje me riprodhimin e aftesive njohese (cognitive) per te automatizuar tasqet. Dallimi eshte Sistemet e AI jane perseritese dhe dinamike duke u bere me inteligjente me sa me shume te dhena qe analizohen.[11]

AI dhe ML meson nga pervoja duke krijuar autonomitetin qe ne te ardhmen vendimet e perseritura te merren ne vendet ku duhet. Nga ana tjeter, analiza e te dhenave (DA) eshte nje proces statik qe shqyrton grupe te medha te te dhenave ne menyre qe te nxjerrim perfundime. DA nuk eshte as perseritje dhe as vete-mesim.[12]

5. Nderthurja e Inteligjences Artificiale me parashikimin e shkeljes se te dhenave.

Pasi tani kemi nje kuptim me te mire te inteligjences artificiale dhe shkeljes se te dhenave, tani duhet edhe te kuptojme impaktin e AI tek parashikimi i shkeljes se te dhenave dhe poashtu ndikimin qe AI posedon tek Siguria Kibernetike. Teknikat e ML dhe “Deep Learning” (DL) do të bëjnë më të lehtë qe te lansohen sulme kibernetike me te sofistikuara, shpejta, fuqishme dhe me shkaterruese.[13] AI ndikon pozitivisht por edhe negativisht nese perdoret per qellime te keqia ne kete fushe. Ndikimi i saj pozitiv e prek shume fushen e parashikimit te shkeljes se te dhenave pasi qe lejon analizen e nje shume te madhe te te dhenave ne kohe shume te shpejte dhe poashtu kupton sjelljet e softuereve te ndryshem dhe eshte ne gjendje qe te ndihmoje ne parandalimin e nje sulmi. Perderisa përdorimi i AI për qëllime mbrojtëse përballet me një numër kufizimesh, siq jane qeveritë (dhe Bashkimi Evropian) lëvizin për të rregulluar aplikacionet me rrezik të lartë dhe për të promovuar përdorimin e përgjegjshëm të AI, në anën tjeter, sulmet vetem shumezohen, kostoja e zhvillimit të aplikacioneve po bie, dhe 'sipërfaqja e sulmit' po bëhet më e dendur çdo ditë, duke e bërë çdo formë të mbrojtjes një betejë ne nje koder te perpjetme.[14] Inxhinieret e sigurise kibernetike qdo dite e me shume mundohen qe te jene ne nje hap te njejte me hakeret te cilet mund te themi se zhvillohen me shpejtesi te dyfishte sa i perket fushes se sigurise kibernetike. ML e ka mundesuar nje detektim shume me te shpejte te sulmeve. Prandaj jane duke u zhvilluar sisteme te reja ku me ane te AI dhe ML te parashikohen sulmet nga hakeret ende pa ndodhur.

---

[11] SISA Editor, (2021), Breach Risk Assessment at *https://www.sisainfosec.com/services/breach-risk-assessment/*

[12] Lbid

[13] Lorenzo Pupillo, Stefano Fantin, Afonso Ferreira, Carolina Polito, (2021, Prill), Artificial Intelligence and cybersecurity, at https://www.ceps.eu/artificial-intelligence-and-cybersecurity/

[14] Lorenzo Pupillo, Stefano Fantin, Afonso Ferreira, Carolina Polito, (2021, Prill), Artificial Intelligence and cybersecurity, at https://www.ceps.eu/artificial-intelligence-and-cybersecurity/

Kjo behet me ane te analitikeve parashikues ku sulmet e ndryshme do parandalohen. Edhe pse 'signatures' metoda per te detektuar sulme potenciale ende eshte nje metode shume e fuqishme, hakeret kane zbuluar exploit kits ku sulmuesi krijon signature unike per qdo lloj sulmi. Ketu mund te shihet fuqia e parashikuesit te shkeljes se te dhenave ku minimizohen false-positive, dhe reagimi do jete shume me i shpejte.[15] Inteligjenca Artificiale dhe parashikimi i shkeljes se te dhenave jo vetem qe mundeson parandalimin e sulmeve, por poashtu parashikohen edhe vendet ku kane dobesi dhe mundesite jane me te medha per sulme, duke dhene informacion ne menyre qe te alokohen masat dhe resurset e duhura mbrojtese.[16]

6.Modeli per parashikimin e rrezikut kibernetik permes analizes se te dhenave dhe 'machine learning'

Eshte e rendesishme te ekzistoje nje model i veqante i cili mund te parashikoje rrezikun e nje nderhyrjeje kibernetike. Menyra se si funksionon modeli eshte duke marrur nje numer te konsiderueshem te te dhenave dhe duke i trajtuar ato me modele te caktuara. Burime te te dhenave ka shume, jane te pafundme, por siq e dime, ne dekaden e fundit koleksionuesit/gjeneruesit me te medhenj te te dhenave kane qene rrjetet e ndryshme sociale si facebook, instagram, twitter etj. Ne do I referohemi nje hulumtimi te vitit 2019 ku jane marre te dhena nga Twitter dhe poashtu te dhena nga CVE Databaza e cila eshte nje database me informacione publike per dobesite e sigurise dhe ekzpozimet ndaj rreziqeve. Datasetet qe perdoren jane dy, njeri data set do jete dataseti I CVE nga viti 1999 deri ne vitin 2017 kurse nga twitteri do perdoret nje dataset njevjeqar, periudha nga viti 2016 ne vitin 2017. Fjalet kyqe per datasetin e twitterit qe perdoren jane: 'vulnerability', 'exploits', 'cve', 'attack', 'zeroday' dhe '0 day' kurse ne menyre qe identifikimi I dobesise te twitterit te perkonte me numrin e identifikimit te dobesise ne datasetin e cve, perdoret fjale kyce 'cve'.[17] Dataseti I CVE nxirret si XML flile kurse dataseti I Twitter nxirret permes API (Nderfaqja e aplikacionit te programit). Databaza e cila eshte perdorur per ti vendosur te dhenat eshte nje MySQL Databaze. Ne kete rast eshte perdorur metoda statistikore e machine learning. Machine learning statistikore perdoret kur nje lidhje statistikore themelohet permes frekuencave te perdorura dhe variables e cila matet pa pasur nevoje te kete lidhje shkakesore e cila mund te jete parametrike, gjysme parametrike, apo jo-parametrike.[18] Lidhja shkakesore ekziston ku nje variable ne nje dataset ndikon direct ne nje variable tjeter. Ne kete punim SML (statistical machine learning) perdoret per bazen e analizave dhe shkon si ne hapat e meposhtem: 1) Identifikimi I problemit, objektivat dhe kerkesat e te dhenave, 2) Koleksionimi I te dhenave, 3) Pastrimi dhe organizimi I te dhenave, 4) Ekstraktimi I vecorive, 5)Trajnimi

[15] Ben Dickson (2016, July), How predictive analytics discovers a data breach before it happens, at *https://techcrunch.com/2016/07/25/how-predictive-analytics-discovers-a-data-breach-before-it-happens/*
[16] SISA Editor, (2021), Breach Risk Assessment at *https://www.sisainfosec.com/services/breach-risk-assessment/*
[17] Athor Subroto, Andri Apriyana, (2019, June), *Cyber risk prediction through social media big data analytics and statistical machine learning*, at *https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0216-1#Sec4*
[18] Global Working Group on Big Data for Official Statistics. Satellite imagery and geo-spatial data. 2017.

dhe testimi I te dhenave me SML, 6) Testimi I saktesise dhe zgjedhja e modelit, 7) Implementimi I modelit. Keta hapa mund ti shohim dhe ne figuren me poshte.

Fig. 2[19]

Pasi qe te happen te dhenat duhet te pastrohen dhe te organizohen dhe pastaj do analizohen me ane ten je analize histogrami, analize wordcloud dhe nje analize dendogrami.[20]

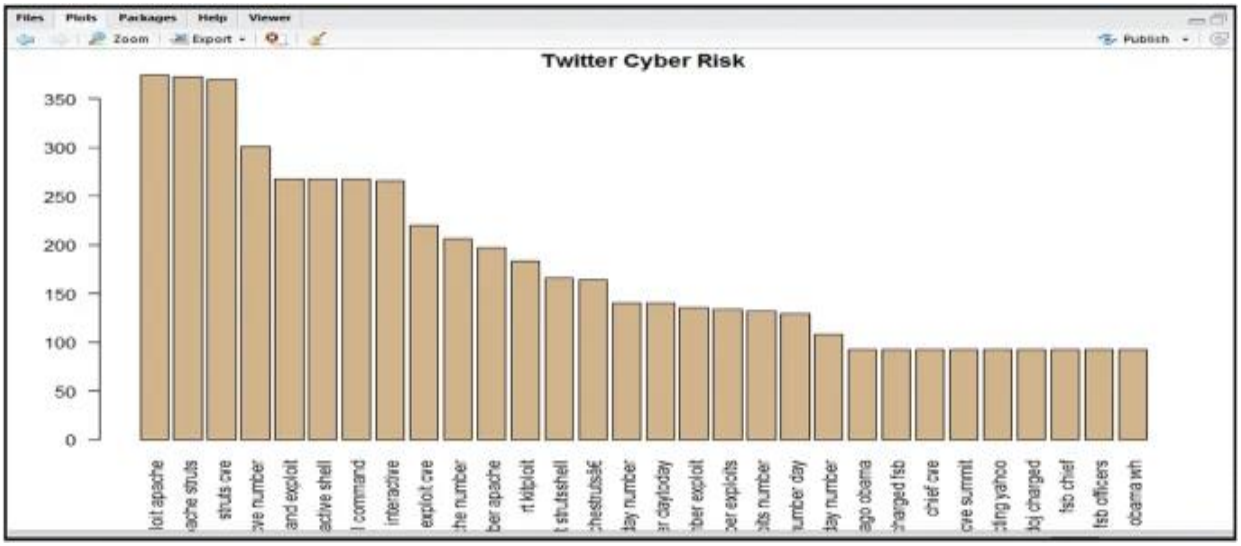


Fig. 3[21]

Me lart shohim analizen e histogramit ku bisedat na tregojne se rreziqet kane ndodhur ne apache, yahoo dhe cisco kurse metodat e sulmit kane qene interactive shell, strutsshell interactive, kitploitstrutshell dhe strutspwn exploit-i.[22]

---

[19] Statistial machine learning modelling phase, at https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0216-1/figures/2

[20] Athor Subroto, Andri Apriyana, (2019, June), *Cyber risk prediction through social media big data analytics and statistical machine learning*, at *https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0216-1#Sec4*

[21] Twitter N-gram histogram analysis.

[22] Athor Subroto, Andri Apriyana, (2019, June), *Cyber risk prediction through social media big data analytics and statistical machine learning*, at *https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0216-1#Sec4*

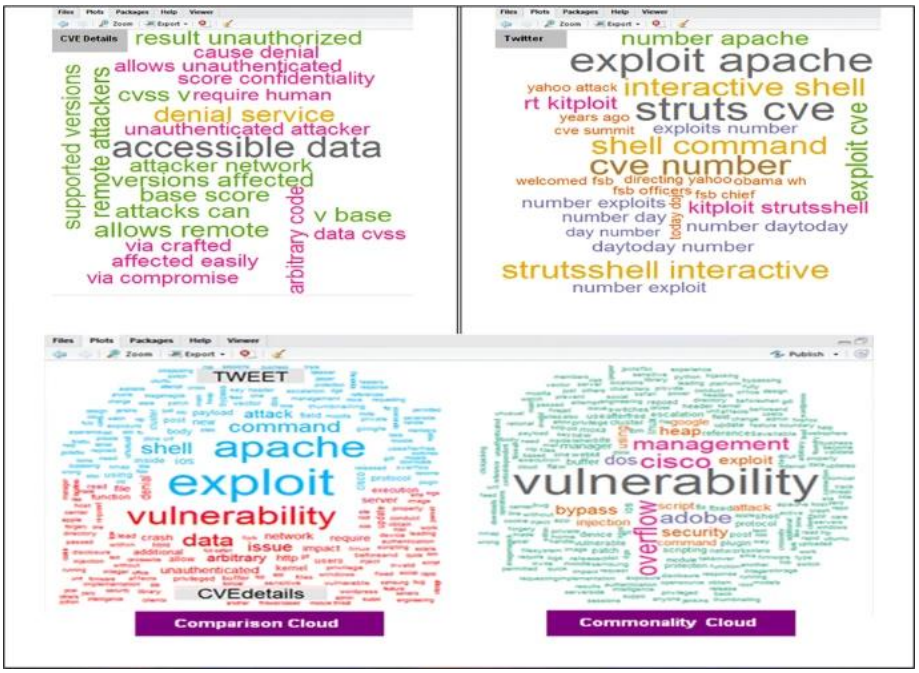


Fig.4[23]

Kjo analize vetem se na ben te kuptojme se analiza e histogramit ka qene e sakte. Apache dhe Struts ka qene bigrami me i frekuentuar.

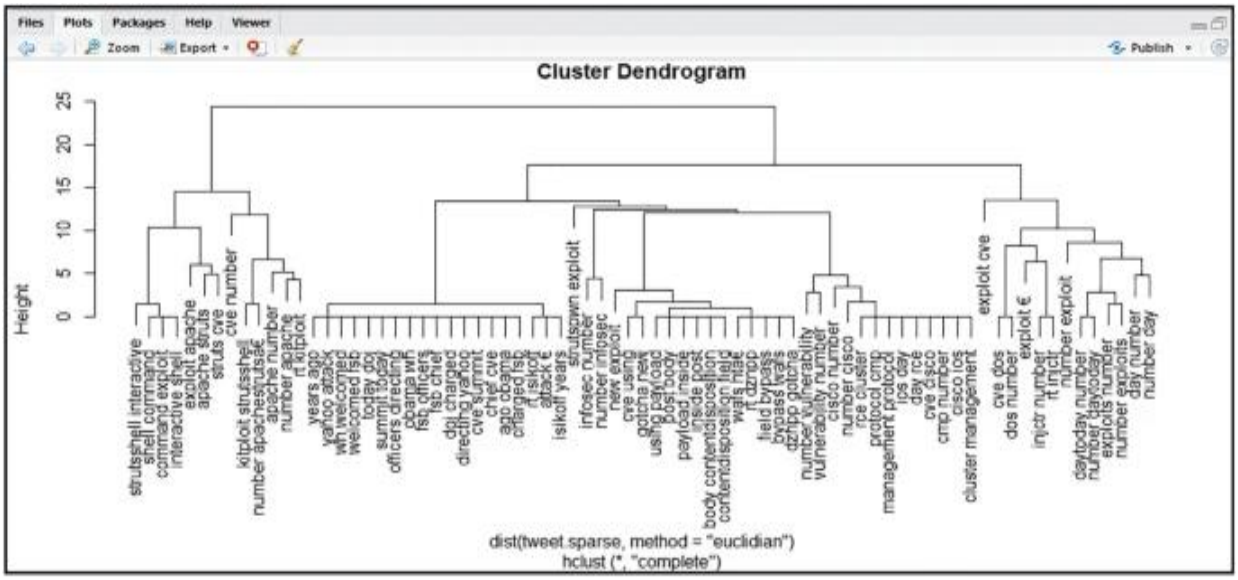


Fig.5[24]

[23] WordCloud analysis, at *https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0216-1#Sec4*

[24] Cluster Dendogram Analysis, at https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0216-1#Sec4

Analiza e cluster dendogramit jep fjalet apache, struts, shell command, exploit dhe interactive. Na ben te dijme se sulmi ne apache struts ka perdorur interactive shell command exploits.[25]

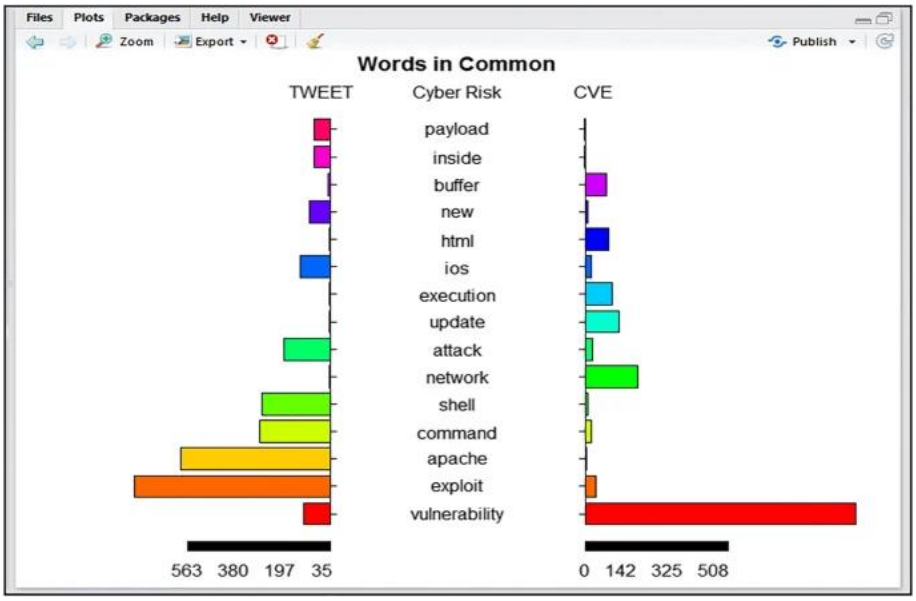
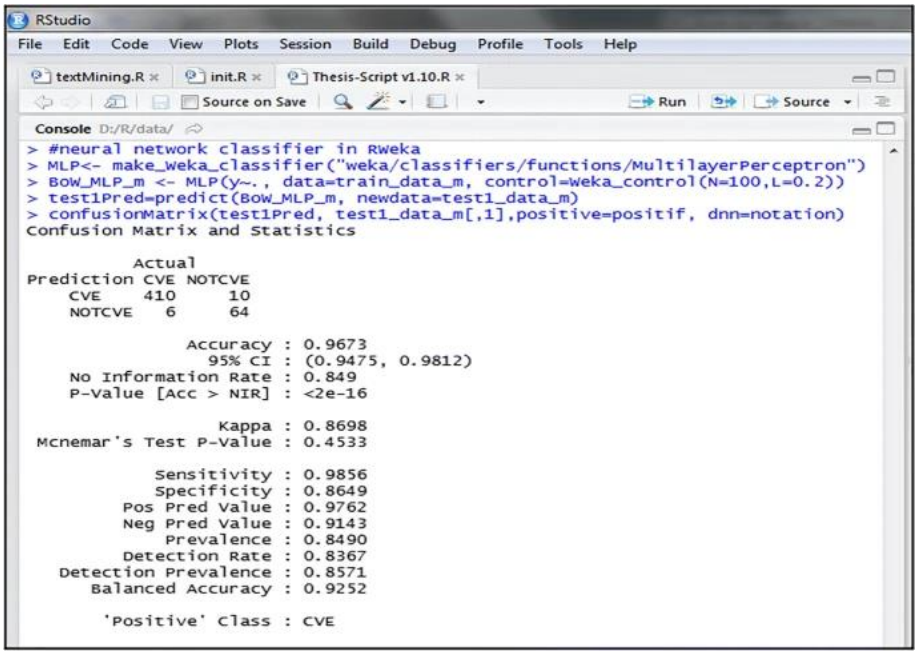


Fig.6[26]

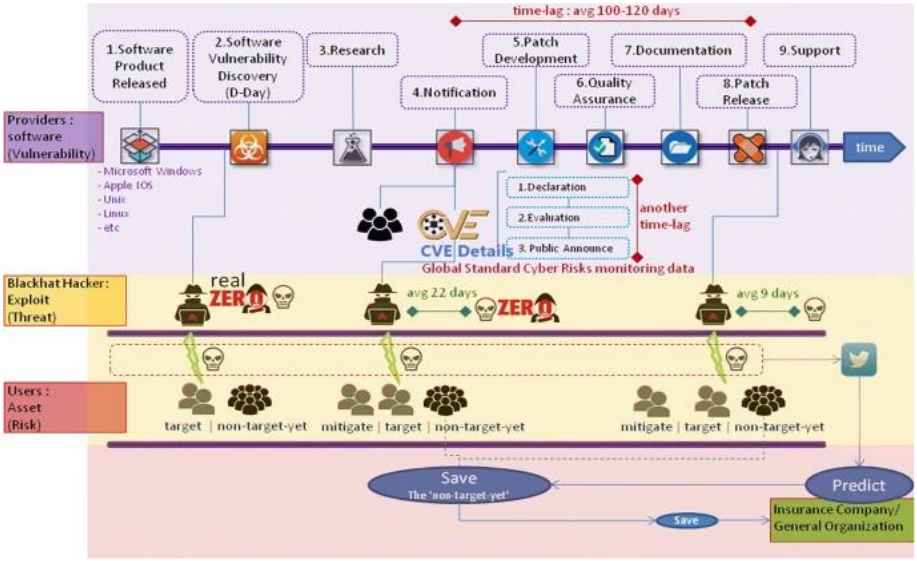
Analiza e piramides apachen e vlereson si unigramin e trete me te frekuentuar, dhe duke u bazuar dhe ne te dhenat paraprake, kuptojme se apache ka rrezikun me te madh per tu sulmuar. Tani kemi arritur tek pjesa e trajnimit te te dhenave dhe testimit permes SML. Per modelin e parashikimit kemi algoritme te ndryshme siq jane: Naive Bayes, Support Vector Machines, Decision Trees, K-nearest neighbors dhe Artificial Neural Networks. Nga te gjitha algoritmet ANN ka rezultatet me te sakta dhe me poshte kemi rezultatet e tij.

[25]Athor Subroto, Andri Apriyana, (2019, June), *Cyber risk prediction through social media big data analytics and statistical machine learning*, at *https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0216-1#Sec4*

[26] Pyramid Analysis, at https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0216-1#Sec4



Ky rezultat parashikon dobesine nga twitter e cila do postohet ne CVE Databaze, duke rezultuar te dihet sa me shpejte per dobesine e re. Per tu llogaritur saktesia eshte perdorur Matrica e konfuzionit. E cila numrin e parashikimeve korrekte dhe jokorrekte e sumarizon me llogaritjen e vlerave duke i ndare secilen ne klase. Perndryshe eshte nje tabele e cila shpjegon performancen e nje modeli klasifikimi ne nje set te te dhenave testuese per te cilen dihen vlerat e verteta.[27] Perndryshe, me algoritmet tjera saktesite kane qene keto: Naive Bayes 55.31%, Super Vector Machine 94.49%m K-nearest neighbor 96.33%, Decision Tree 94.08% dhe Artificial Neural Network me 96.73%. Ne praktike implikimi i modelit do ishte si ne fifuren me poshte ku secili hap duhet te percillet me rend.



[27] Data School Author, (2014, March), Simple Guide to confusion matrix terminology, at https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/

Fig.8[28]

Kuptojme se pasi te lirohet produkti ne market, ka mundesi qe hakeret te. Zbulojne dobesi dhe te sulmojne pedorues te ndryshem qe mund te jene perdorues shenjestra apo perdorues te rendomte. Pasi qe te kuptohet se ekziston nje dobesi e tille, eshte vetem qeshtje kohe qe edhe hakeret te zbulojne kete dhe te rrisin intensitetin e sulmeve te tyre ku ne poashtu shohim dhe kohen qe u nevojitet kompanise qe ta mbulojne dhe rregullojne problemin. Pra teknikisht, eshte nje lloj lufte kohe kunder hakereve ne menyre qe sistemi te perditesohet sa me shpesh te jete e mundshme dhe te minimizohet rreziku i nderhyrjes nga hakeret.

7.Konkluzioni

U kuptua qe vetem vitet e fundit ka pasur rritje drastike ne numrin e sulmeve por poashtu ka pasur avancim ne pjesen e sigurise per shkak te implementimit te AI dhe ML ne sistemet e sigurise. AI ne siguri kibernetike ka ndikuar shume duke marre ne dore pune te cilat nuk do kishin mundesi te kryhen nga nje person apo edhe nga nje numer i madh i personave. AI mund ta quajme si nje shpate me dy tehe, per shkak se me avancim te mbrojtjes dhe lehtesim te parandalimit te sulmeve, avancohen edhe akteret kercenues duke shumefishuar dhe nderruar taktikat e tyre sulmuese duke perdorur te njejten teknologji. AI ka arritur deri ne ate pike saqe nje tentim shkelje e te dhenave mund te detektohet ende pa pasur shkeljen. Nga te gjitha keto informacione, na ben te kuptojme se buxheti sa i perket mbrojtjes me AI eshte shume i vogel ne shume entitete te ndryshme. Si zgjidhje e ketij problemi do ishte vetedijesimi i entiteteve perkatese sa i perket sigurise, nje rregullore nderkombetare per implementimin e AI ne sistemet e sigurise dhe bashkepunimi mes institucioneve publike dhe private. Poashtu kuptuam nga modeli se cilat mund te jene pikat qe mund te targetohen nga sulmuesit me se shumti dhe me ane te parashikimeve permes algoritmeve kuptuam se me ane te ANN ne kishim saktesi te larte prej 96.73% dhe me ane te ketij modeli ne mund te parashikojme se cilat nga sherbimet ne nje kompani te caktuar ka potencial me se shumti te jete caku i ndonje sulmi nga hakeret. Pra, modeli ne fjale mund te jete nje model shume efektiv per inxhinieret e sigurise te krijojne strategji efektive per te ulur rreziqet potenciale ne sistemin e tyre dhe poashtu te identifikojne pikat te cilat duhet te forcohen me shume.

[28] Cyber Risk Occurrence Figure, at https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0216-1#Sec4