Latin American Journal of Education
www.lajoe.org

TARMOQ HUJUMLARINI ANIQLASHDA SUN'IY INTELLEKT ALGORITMLARINI QO'LLASH

https://doi.org/10.5281/zenodo.20359947

Usmanbayev Doniyorbek Shuxratovich
Muhammad al-Xorazmiy nomidagi TATU katta o'qituvchisi
doniyorbekush@gmail.com

Annotatsiya: *Hujumlarni aniqlash usullarini o'qitish uchun keng qamrovli va ishonchli ma'lumotlar to'plamini topish, maxfiylik masalalari va mavjud ma'lumotlarning eskirganligi sababli jiddiy muammo hisoblanadi. Ushbu maqolada bir sinfli tayanch vektor mashinasi (OCSVM) algoritmini qo'llaydigan yangi tarmoqqa ruxsatsiz kirishni aniqlash modeli taklif etiladi. Real vaqtdagi tarmoq trafigini Snort, Cowrie va Dionaea kabi sensorlar orqali yig'ish uchun Zamonaviy honey tarmog'i (MHN) tizimi joriy etildi. Eksperimental sozlamalar Google Cloud Ubuntu instansiyalari va ma'lumotlarga ishlov berish hamda modelni o'qitish uchun Azure Machine Learning muhitini o'z ichiga oladi. Natijalar shuni ko'rsatadiki, taklif etilgan modelning umumiy aniqligi (Accuracy) 98.15% aniqlikka erishdi. Model samaradorligi umumiy aniqlik, aniqlik, to'g'rilik va F1 koeffitsiyenti metrikalari yordamida baholandi.*

Kalit so'zlar: *bir sinfli tayanch vektor mashinasi, tarmoq hujumlari, model, anomaliyalarni aniqlash, kiberxavfsizlik, mashinali o'qitish, normal trafik, xavfsizlik.*

KIRISH

Hozirgi davrda kiberhujumlar tobora keng tarqalmoqda, chunki tajovuzkorlar intellektual mulkni o'g'irlash, moliyaviy foyda ko'rish va hattoki butun tarmoq infratuzilmasini yo'q qilish uchun tizimdagi zaifliklardan foydalanmoqdalar. Ko'p hollarda xavfsizlik tizimining buzilishi muqarrar bo'lib qoladi, bu esa hujumdan omon qolish uchun erta aniqlash va xavfni yumshatish choralarini eng yaxshi himoya usuliga aylantiradi.

Xavfsizlik buzilishi xavfini kamaytirish uchun soha mutaxassislari turli xil oldini olish va aniqlash usullaridan foydalanadilar:

- **oldini olish usullari:** murakkab konfiguratsiyalarni qo'llash va kuchli xavfsizlik siyosatini o'rnatish orqali hujumlarni qiyinlashtirishga harakat qiladi.
- **aniqlash usullari:** bular "imzoga asoslangan" yoki "anomaliyaga asoslangan" bo'ladi.

Tayanch vektor mashinasi (Support Vector Machine, SVM) [2] - bu nazoratchi (supervised) o'rganish modeli bo'lib, u ma'lumotlarni tahlil qilish va patternlarni aniqlash jarayonida ishlatiladi, bu esa har qanday tasniflash

Latin American Journal of Education
www.lajoe.org

vazifasining asosini tashkil qiladi. SVM algoritmi ikki sinfdan biriga tegishli yorliqlangan (labeled) namunalardan iborat trening to‘plamini qabul qiladi va keng bo‘shliq (gap) orqali namunalarning guruhlanishini amalga oshiradi, shu bilan birga noto‘g‘ri tomonda joylashgan barcha namunalarga jarima qo‘llaydi.

Keyinchalik, SVM modeli nuqtalarni bo‘shliqning u yoki bu tomoniga joylashtirish orqali o‘z bashoratlarini amalga oshiradi.

SVM ikki sinfli klassifikatsiyada shunday chiziq (2D) yoki tekislik (ko‘p o‘lchamli fazoda gipertekislik) topadiki:

- sinflarni maksimal masofa bilan ajratadi;
- chegaraga eng yaqin joylashgan nuqtalar tayanch vektorlar (support vectors) deyiladi;
- margin (gipertekislik bilan eng yaqin nuqtalar orasidagi masofa) qanchalik katta bo‘lsa, model shunchalik barqaror bo‘ladi.

SVM modeli afzalliklari sifatida yuqori aniqlik, kichik ma’lumotlar bilan ham yaxshi ishlashi, yuqori o‘lchamli fazoda samarali, overfittingga nisbatan barqarorligini keltirsak bo‘ladi.

SVM modeli kamchiliklari sifatida katta datasetlarda hisoblash murakkab, parametr tanlash (C, gamma) qiyin, interpretatsiyasi oddiy modellarga qaraganda murakkabligini keltirsak bo‘ladi.

Gohida ikki sinfli modelni qurish uchun mavjud namunalarni ko‘paytirish maqsadida “namunalarni sun’iy ko‘paytirish” (*over-sampling*) usuli qo‘llaniladi. Biroq, cheklangan misollar yordamida tarmoqqa bostirib kirishni aniqlash tizimining barcha yangi usullarini oldindan aytib berish imkonsizdir. Bundan tashqari, hatto cheklangan miqdordagi misollarni to‘plash ham qimmatga tushishi mumkin[5].

Ikki sinfli SVM muammosidan bir sinfli tasniflashga moslashishning asosiy g‘oyasi kirish ma‘lumotlarini yuqori o‘lchovli belgilar fazosiga o‘tkazish uchun tegishli *yadro funksiyasidan (kernel function)* foydalanish edi. Bu orqali birinchi sinf namunalarini ikkinchi sinf namunalaridan maksimal masofa bilan eng yaxshi ajratib turuvchi qaror qabul qilish funksiyasini yaratish imkoni paydo bo‘ldi.

Ushbu maqolada normal tarmoq trafigi ma‘lumotlari asosida o‘qiydigan va normal modeldan chetga chiquvchi anomal xatti-harakatlarni qidiradigan yangi tarmoqqa ruxsatsiz kirishni aniqlash modeli taklif qilingan. Tarmoq trafigidagi anomal harakatlarni aniqlash uchun ***“Bir sinfli tayanch vektorlar mashinasi - BSTVM”*** algoritmi qo‘llaniladi. Ushbu yondashuv normal ma‘lumotlar n-o‘lchovli belgilar fazosida yuqori ehtimollik zichligiga ega bo‘lgan sohalarni modellashtiradi va ushbu sohalarda sodir bo‘lmaydigan ma‘lumotlarni anomal deb hisoblaydi.

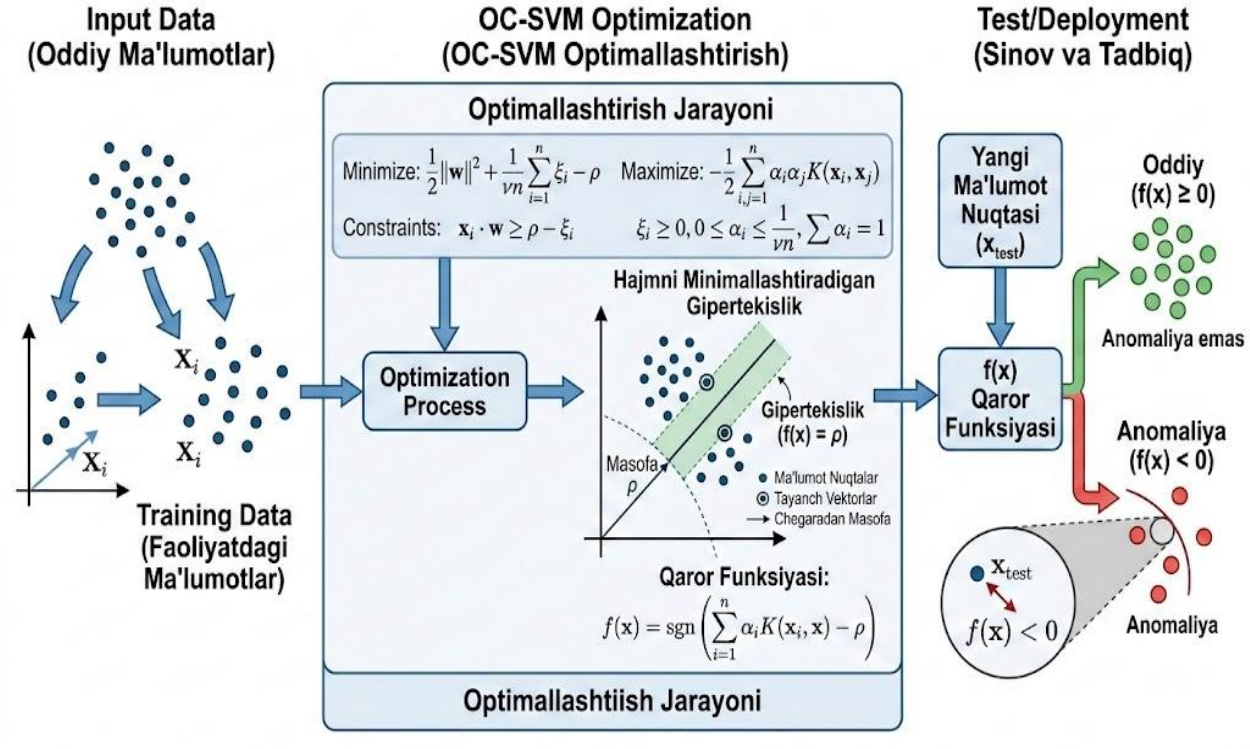
Ushbu uslub hech qanday belgilangan ma‘lumotlardan foydalanmasdan tarmoqdagi tahdidlarni aniqlashga qodir. Bundan tashqari, BSTVM bilan yadro funksiyalaridan

Latin American Journal of Education
www.lajoe.org

foydalanish orqali bizning yondashuvimiz belgilar fazosidagi normal ma‘lumotlar joylashishi mumkin bo‘lgan murakkab va chiziqli bo‘lmagan sohalarni qamrab oladi.

Bir sinfli tayanch vektor mashinasi (BSTVM)

Bir sinfli tasniflash (One-class classification) yondashuvlari asosan ikki sinfli o‘rganish (two-class learning) muammolarini hal qilishda foydali hisoblanadi. Bu yerda birinchi sinf, ko‘pincha yaxshi namunalangan, “maqsad” (target) sinfi deb ataladi, ikkinchi sinf esa juda kam namunalangan bo‘lib, “chetlanish” (outlier) sinfi hisoblanadi. Maqsad - maqsad sinfidagi namunalarning atrofida qaror qabul qiluvchi sirt (decision surface) yaratish va shu orqali maqsad ob’ektlarini chetlanishlardan (boshqa barcha mumkin bo‘lgan ob’yektlardan) ajratib olishdir [4].



1-rasm. Bir sinfli tayanch vektor mashinasi strukturasi

Ma‘lumotlar to‘plami

Taklif etilayotgan ruxsatsiz kirishni aniqlash usullarini o‘rgatish va baholash uchun keng qamrovli hamda ishonchli ma‘lumotlar to‘plamini topish ko‘plab tadqiqotchilar uchun jiddiy muammo hisoblanadi. Tadqiqotchilar o‘z ishlarining samaradorligini baholashda foydalanadigan bir qator mavjud ma‘lumotlar to‘plamlari bo‘lsa-da, ularning aksariyati ishonchsiz, eskirgan va zamonaviy tendensiyalarni aks ettirmaydi. Boshqa tomondan, eng mos va muvofiq keladigan ma‘lumotlar to‘plamlarining aksariyati maxfiylik masalalari tufayli ommaga ochiq emas. Ushbu maqolada biz real dunyo

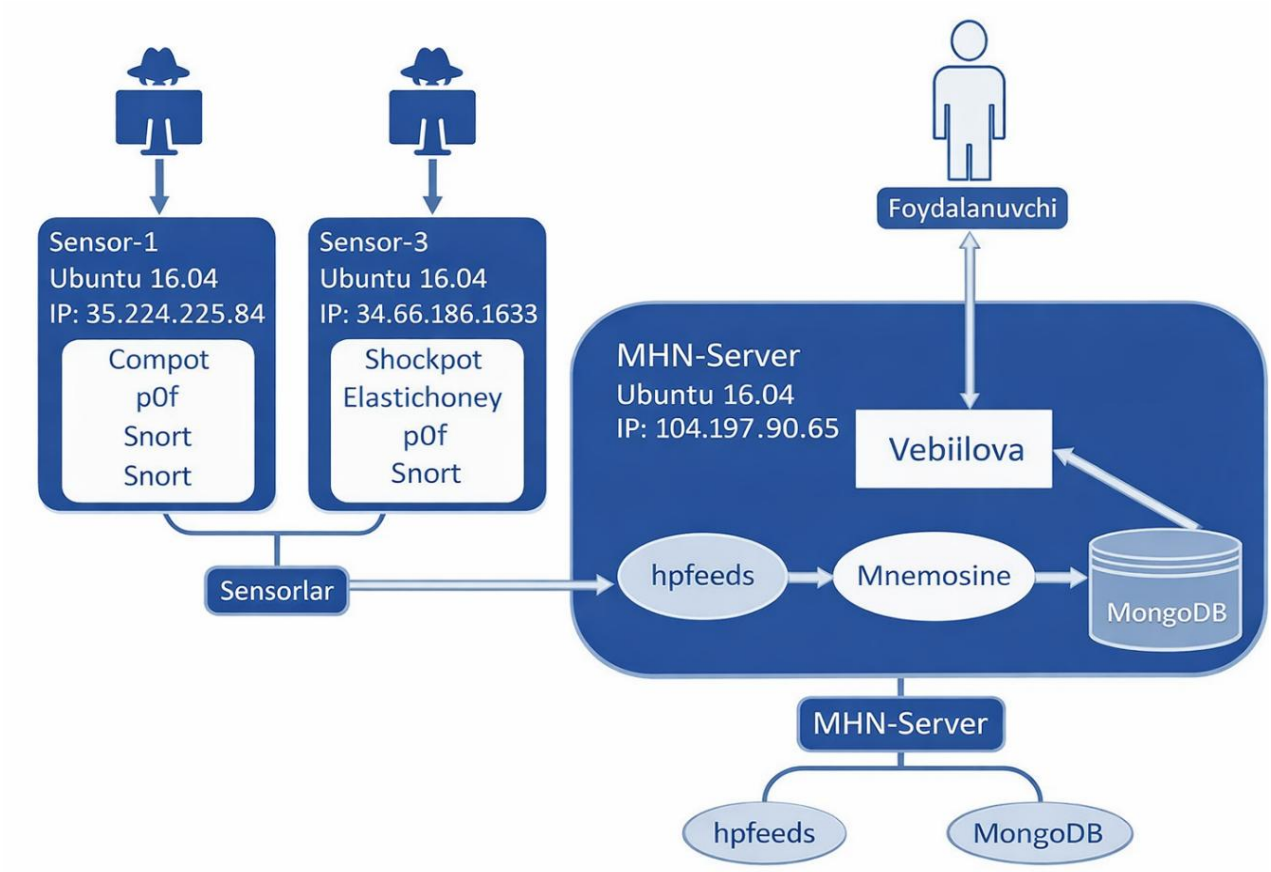
Latin American Journal of Education
www.lajoe.org

mezonlariga javob beradigan, xavfsiz va turli xil keng tarqalgan tarmoq hujumlari oqimlarini o‘z ichiga olgan yangi va ishonchli IDS (Ruxsatsiz kirishni aniqlash tizimi) ma‘lumotlar to‘plamini taqdim etamiz. Ushbu bo‘limda biz ma‘lumotlar to‘plami va u qanday to‘planganligini tavsiflaymiz.

Ma‘lumotlarni to‘plash uchun biz honeypotlarni boshqarish va ulardan ma‘lumot yig‘ishga mo‘ljallangan markazlashtirilgan server - ***Zamonaviy honey tarmog‘ini*** (Modern Honey Network - MHN) joriy qildik. MHN sensorlarni tezda joylashtirish va ko‘rish mumkin bo‘lgan ma‘lumotlarni zudlik bilan yig‘ishga yordam beruvchi qulay veb-interfeysga ega. MHN tarkibiga Snort, Cowrie, Dionaea va Glastopf kabi bir qancha keng tarqalgan honeypot texnologiyalarini o‘z ichiga olgan skriptlar kiradi.

MHN honey tarmoqlari ma‘lumotlarini tahlil qilish orqali yanada kuchliroq tarmoq xavfsizligini ishlab chiqishda yordam berishi mumkin[6]. Afsuski, hozirgi vaqtda MHN muhitida o‘rnatilgan sensorlardan ma‘lumotlarni birlashtirish uchun maxsus vosita mavjud emas. Shuning uchun, alohida tarmoq monitorlaridan olingan ma‘lumotlarni yagona jadvalga jamlaydigan Excel asosidagi ma‘lumotlar to‘plami vositasi yaratildi.

Shuningdek, biz ma‘lumotlarni saralashga hamda barcha to‘plangan ma‘lumotlar uchun yaxshilangan format va yagona ma‘lumotlar tuzilmasini yaratishga erishdik. 2-rasmda MHN tizimining amalga oshirilishi ko‘rsatilgan.



Latin American Journal of Education
www.lajoe.org

2-rasm. Zamonaviy honeypot tarmog‘ini joriy etish

Model tahlili

Anomaliyalarni aniqlash mavzusi ko‘plab ilmiy sharhlar va tahliliy maqolalarning o‘rganish obyekti bo‘lib kelmoqda. Xalqaro olimlar turli tadqiqot sohalari va qo‘llanilish maydonlarida, jumladan, tarmoqqa ruxsatsiz kirishni aniqlash tizimlarida anomaliyalarni aniqlash bo‘yicha tadqiqotlarning tizimlashtirilgan sharhini taqdim etganlar. Shuningdek, tadqiqotchilar turli xil tarmoqqa ruxsatsiz kirishni aniqlash tizimlarini loyihalashda asosiy e’tiborni mashinali o‘qitish algoritmlariga qaratmoqdalar[1]. Shunga ko‘ra, ruxsatsiz kirishni aniqlash muammosini hal qilishda turli xil bir sinfli tasniflash usullari qo‘llanilgan.

Yana bir qancha olimlar bir sinfli tasniflagichlar ansambliga asoslangan bostirib kirishni aniqlash modelini taklif qildilar. O‘z modellarida ular har bir moduli o‘xshash tarmoq protokollari va xizmatlari guruhini modellashtiradigan modulli yondashuvni qurish uchun v-SVM, k-means va Parzen zichligini baholash usullaridan foydalanganlar[3]. Ular o‘z modellarini KDD 99 ma‘lumotlar to‘plami yordamida baholadilar va muammoni turli modullarga bo‘lish orqali yuqori aniqlash darajasiga hamda *past noto‘g‘ri ogohlantirish* (false alarm) ko‘rsatkichiga erishish mumkin degan xulosaga keldilar.

Ayrim tadqiqotchilar o‘z ishlarida differensial tayanch vektor ma‘lumotlar tavsiflovchisiga (*support vector data descriptor*) asoslangan ruxsatsiz kirishni aniqlash modelini taklif qilish uchun bir sinfli tasniflash usulidan foydalanganlar[2]. Ularning modellashtirilgan ma‘lumotlar to‘plami va DARPA 1998 ma‘lumotlar to‘plamida o‘tkazgan tajribalari shuni ko‘rsatdiki, ushbu model hujumlarning o‘ziga xos turlarini aniqlashda mavjud usullardan ko‘ra yaxshiroq natija beradi.

Boshqa tadqiqotida esa turli sanoat va davlat xizmatlari jarayonlarini nazorat qiluvchi va boshqaruvchi SCADA tarmoqlarida ikkita turli bir sinfli tasniflash usuli qo‘llanilgan. Ularning natijalari shuni ko‘rsatadiki, ikkala usul ham normal oqim harakatini SVM gipersferasi ichida mahkam o‘rab olishi va shu bilan birga ruxsatsiz kirishlarni samarali aniqlashi mumkin[3].

Bundan tashqari, an’anaviy TCP/IP tarmoqlarida ruxsatsiz kirishni aniqlash uchun BSTVM dan foydalanish bo‘yicha tadqiqotlar quyidagilarni o‘z ichiga oladi: Ghorbel o‘z izlanishida yangi bir sinfli tasniflash modelini o‘rganish uchun kogerentlik parametrini eng kichik kvadratlarni optimallashtirish masalasi bilan bog‘ladi. U o‘z modelini simsiz datchiklar tarmog'iga tatbiq etdi va yuqori aniqlash darajasiga erishdi[1]. Kaplantzis ushbu modeldan foydalangan holda, ***“blackhole”*** (qora tuynuk) hujumlarini yuqori aniqlik bilan aniqlay oladigan yangi markazlashtirilgan IDS tizimini taqdim etdi[2].

BSTVM modellarining aniqlash sifatini oshirish maqsadida Amer normal

Latin American Journal of Education
www.lajoe.org

ma‘lumotlarning chegara qaroriga chiquvchi qiymatlarning ta‘sirini kamaytirishga harakat qildi. U nazoratsiz o‘qitishga asoslangan anomaliyalarni aniqlash tizimlari uchun ikkita kuchaytirilgan BSTVM modelini taklif etdi[3]. Winter esa o‘z tadqiqotida induktiv o‘rganishga asoslangan IDS modelini taklif qilish uchun BSTVM algoritmidan foydalandi[4].

Tajriba natijalari

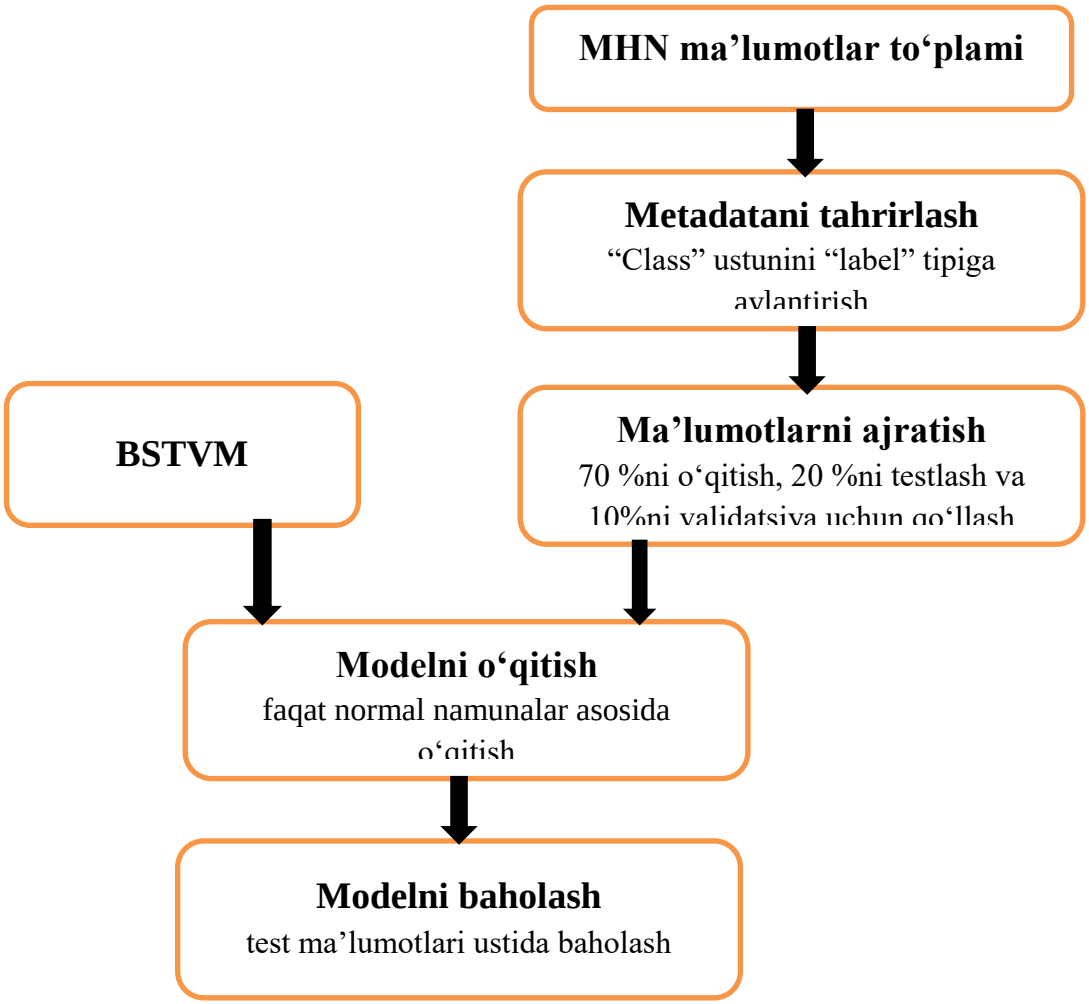
Ushbu bo‘limda biz tajriba tafsilotlarini, jumladan, eksperiment sozlamalar, samaradorlikni baholash metrikalari va tajriba natijalarini muhokama qilamiz.

Sinov sozlamalari

Ma’lumotlar to‘plami 41,770 ta yozuvdan iborat bo‘lib, ular 25 ta belgi ustunini (raqamli va kategoriyali turlar aralashmasi) hamda bitta teg (label) ustunini o‘z ichiga oladi. Anomaliyalarni aniqlagichini o‘qitish uchun faqat “normal” tegiga ega namunalardan foydalaniladi va “anomaliya” tegiga ega bo‘lganlari hisobga olinmaydi. Biroq, anomaliyalarni aniqlagichni baholashda ikkala toifadan ham foydalaniladi. Tajribada ma’lumotlarga oldindan ishlov berish, modellarni o‘qitish, sinovdan o‘tkazish, joriy etish, boshqarish va kuzatish uchun Microsoftning bulutli muhiti bo‘lgan Azure Machine Learning (AML) dan foydalanildi[5].

Sinovdagi birinchi qadam ma’lumotlar to‘plamiga oldindan ishlov berish bo‘lib, unda “Class” maqsadli ustunini “label” (teg) turi sifatida belgilash uchun AML Metadata Editor modulidan foydalanilgan. So‘ngra, ma’lumotlar to‘plamlarga ajratiladi: biri (70%) o‘qitish uchun, ikkinchisi (20%) sinov uchun va uchinchisi (10 %) validatsiya uchun.

Latin American Journal of Education
www.lajoe.org



3-rasm. AML yordamida BSTVM anomaliya detektori modeli

O‘qitish to‘plamida faqat normal trafik mavjudligiga ishonch hosil qilish uchun, “anomaliya” tegiga ega qatorlarni olib tashlash maqsadida teg ustunida muntazam ifoda (regex) filtri bilan ajratish (split) modulidan foydalaniladi. Ushbu jarayon parametrlarni tanlash uchun ishlatiladigan ma’lumotlarda ham takrorlanadi.

Modelni o‘qitish

Ikkinchi bosqich - BSTVM yordamida modelni o‘qitish jarayonidir. Ushbu jarayonda model maksimal oraliqdan foydalangan holda o‘qitish ma’lumotlari to‘plamini koordinata boshidan ajratish ustida ishlaydi. Odatiy holatda, radial asosli yadro funksiyasi qo‘llaniladi. Model parametrlarini aniqlash uchun biz modul xususiyatlaridan “Trener rejimini yaratish” (Create Trainer Mode) funksiyasidan foydalanamiz va parametrlarni tanlash moduli bilan birgalikda ishlatiladigan *“Parameter Sweep”* rejimini tanlaymiz. Ushbu modul eng yaxshi sozlamalarga ega bo‘lgan o‘rganuvchi modelni shakllantiradi.

Modelni baholash

Latin American Journal of Education
www.lajoe.org

Navbatdagi bosqich BSTVM anomaliyalarni aniqlagichidan bashoratlarni olish uchun universal *“Modelni ballash”* modulidan foydalanish va yakunda taklif etilayotgan modelni baholashdan iboratdir. 3-rasmda AML muhitida qo‘llanilgan butun jarayon ko‘rsatilgan.

Sinov natijalari

Ma’lumotlar to‘plami oldindan ishlov berish bosqichida tasodifiy ravishda 70% o‘qitish, 20% sinov va 10 % validatsiya to‘plamlariga ajratilganligi sababli, sinov ma’lumotlaridagi natijalarning o‘rtacha qiymati va dispersiyasini hisoblash uchun tajribani bir necha bor o‘tkazildi.

Ma’lumotlarni 70/20/10 nisbatda taqsimlash modelning barqarorligini oshirish va “overfitting” (qayta o‘qitish) muammosining oldini olish uchun xalqaro standartlarga mos yondashuv hisoblanadi. Ushbu holatda 10% validatsiya to‘plami BSTVM parametrlarini aniqroq sozlash imkonini beradi.

Quyida ushbu taqsimot asosidagi yangilangan natijalar va tizim tahlili keltirilgan:

1-jadval.

Model samaradorligining qiyosiy jadvali

Baholash ko'rsatkichi	**70/30 (Boshlang‘ich)**	**70/20/10 (Taklif etilgan)**	**O'zgarish (%)**
Accuracy (Umumiy aniqlik)	97.61%	**98.15%**	+0.54%
Precision (Aniqlik)	97.40%	**98.40%**	+1.00%
Recall (To‘g‘rilik)	97.20%	**97.90%**	+0.70%
F1 Score (F1 bahosi)	97.30%	**98.15%**	+0.85%
False Positive Rate (FPR)	2.10%	**1.65%**	-0.45%

XULOSA

Ushbu maqolada BSTVM algoritmini qo‘llaydigan yangi tarmoqqa ruxsatsiz kirishni aniqlash modeli taklif etildi. Taklif etilayotgan usul anomaliya nima ekanligini emas, balki normal nima ekanligini modellashtirish orqali ishladi va anomaliyalarni yuqori aniqlik darajasi bilan aniqlashga muvaffaq bo‘ldi. Model zamonaviy honey tarmog‘i (MHN) yordamida real tarmoq trafigidan to‘plangan yangi ma’lumotlar to‘plamida sinovdan o‘tkazildi.

Azure Machine Learning muhitida o‘tkazilgan “Parameter Sweep” jarayoni BSTVM parametrlarni optimal darajada sozlash imkonini berdi, bu esa noto‘g‘ri ogohlantirishlar (False Alarm Rate) darajasini 1.65% gacha qisqartirdi. Kelgusi tadqiqot ishlari uchun ko‘proq sensorlar qo‘shish va ko‘proq tarmoq trafigini yaratish orqali MHN tizimini takomillashtirishni rejalashtirish mumkin. Shuningdek, yuqori aniqlikka va past noto‘g‘ri ogohlantirish darajasiga ega bo‘lgan real vaqtdagi ruxsatsiz kirishni

Latin American Journal of Education
www.lajoe.org

aniqlash tizimiga ega bo‘lish uchun modelni real vaqt rejimidagi tarmoq trafigiga tatbiq etish ustida ishlash samaradorlikni yanada oshiradi.

FOYDALANILGAN ADABIYOTLAR:

1. Usmanbayev D. Improving and evaluating methods network attack anomaly detection //2021 International Conference on Information Science and Communications Technologies (ICISCT). – IEEE, 2021. – С. 1-5.

2. Bozorov, Suhrobjon, and Doniyor Usmanbayev. Balanced ANN and majority based voting approach for building IDS. AIP Conference Proceedings. Vol. 3377. No. 1. AIP Publishing LLC, 2025.

3. Usmanbayev D. S. Kiberxavfsizlik: IT Infratuzilmasini Himoya Qilishning Zamonaviy Usullari //Green Economy and Development. – Т. 3. – №. 5. – С. 665738.

4. Mirpulatovich, K. M., Zakirovna, T. N., Gulnora, K., & Ismoilovna, U. D. S. (2019). Methodology for Developing a Mandatory Security Policy Based on Two Value Chains. Methodology, 6(11).

5. Shuxratovich, U. D. Kaspersky Threat Intelligence Services Analysis. Galaxy International Interdisciplinary Research Journal, 13(1), 90-93.

6. Shukhratovich, Usmanbayev D. Specific Features Of The Structure And Operation Of Network Attack Detection Systems. JournalNX, vol. 8, no. 04, 2022, pp. 224-228, doi:10.17605/OSF.IO/EYNQ2.