СОӨЖ №1. Кибершабуылдарды анықтауда AI/ML қолдану бағыттарын талдау

Жұмыстың мақсаты

Кибершабуылдарды анықтауда AI/ML қолданылатын негізгі бағыттарды талдау, әр міндетке сәйкес модель түрін, деректерді және бағалау критерийлерін негіздеу, семестрлік жоба үшін перспективалы AI/ML сценарийін таңдау.

Пәнмен және жеке семестрлік жобамен байланысы

СОӨЖ №1 СӨЖ №1-де қарастырылған аномалиялық анықтауды тереңдетеді және ПЖ №5–7-дегі preprocessing, ML классификациясы және Suricata+ML гибридті архитектурасына дайындық жасайды. Жұмыс нәтижесі семестрлік жобаның ML компонентіне қойылатын талаптарды қалыптастырады.

Қысқаша теориялық мәліметтер

AI/ML киберқауіпсіздікте үлкен көлемдегі оқиғалардан қайталанатын заңдылықтарды табу, аномалияларды анықтау, оқиғаларды жіктеу және талдаушыға басымдық беруге көмектесу үшін қолданылады. ML қолдану мақсаты міндетті түрде «адамды алмастыру» емес; көбіне ол көп көлемдегі деректерді алдын ала сүзу және күмәнді оқиғаларды бөліп көрсету құралы болып табылады.

Кибершабуылдарды анықтауда supervised learning қалыпты және шабуыл кластары алдын ала белгіленген деректермен жұмыс істейді. Classification міндетінде Logistic Regression, Decision Tree, Random Forest, Gradient Boosting және нейрондық желілер сияқты модельдер қолданылуы мүмкін. Модель таңдау тек күрделілікке емес, деректер көлеміне, белгі түріне және түсіндірмелілік талабына байланысты.

Unsupervised және semi-supervised тәсілдер белгіленген шабуыл мысалдары жеткіліксіз болған жағдайда пайдалы. Олар қалыпты трафиктің құрылымын немесе деректердегі топтарды зерттеп, одан ауытқитын нүктелерді аномалия ретінде белгілеуі мүмкін. Мұндай нәтижелерде False Positive тәуекелі жоғары болуы мүмкін, сондықтан контекстік тексеру қажет.

Желілік қауіпсіздікте AI/ML ағындық трафикті жіктеу, ботнет немесе DDoS белгілерін анықтау, аномалиялық DNS белсенділігін зерттеу және IDS alert-терін басымдықтау үшін қолданылуы мүмкін. Endpoint немесе журнал деректерінде пайдаланушы әрекеті, процесс оқиғалары және аутентификация үлгілері зерттеледі. Әр міндет үшін feature space пен label мағынасы әртүрлі болады.

Модельдің сапасы dataset сапасына тікелей тәуелді. Класстар теңгерімсіздігі, data leakage, қайталанған жазбалар, бір шабуыл кампаниясының train және test жиындарына араласуы нәтижені жасанды жоғарылатуы мүмкін. Сондықтан preprocessing және split стратегиясы модель архитектурасынан кем емес маңызды.

AI/ML жүйесін бағалауда Precision, Recall, F1-score және confusion matrix негізгі метрикалар болып табылады. Қауіпсіздік міндетінде False Negative нақты қауіпті өткізіп алу тәуекелін білдірсе, False Positive талдаушыға артық alert жүктемесін береді. Threshold таңдау осы екі қатенің арасындағы теңгерімге әсер етеді.

Магистрант үшін маңызды міндет – ең күрделі модельді емес, зерттеу сұрағына сәйкес келетін модельді таңдау. Baseline қарапайым әрі түсіндірілетін бастапқы нәтиже береді, ал күрделі модельдің артықшылығы дәл сол шарттарда өлшенуі тиіс. Семестрлік жобада модельдің дәлдігімен қатар latency, reproducibility және түсіндірмелілік те ескеріледі.

Әдістемелік нұсқаулар

Оқытушымен бірге семестрлік жобадағы қауіпсіздік міндетін нақтылаңыз: классификация, аномалияны анықтау, alert prioritization немесе басқа қорғаныстық сценарий. Осы міндетке сәйкес кіріс деректердің түрін және target/label мағынасын анықтаңыз.

Кемінде үш AI/ML тәсілін таңдаңыз: бір baseline және екі балама. Әр тәсіл үшін қажетті preprocessing, feature түрлері, артықшылықтары, шектеулері және түсіндірмелілік деңгейін салыстырыңыз.

Пайдаланылатын ашық dataset немесе оқытушы берген деректердің шығу тегін, класстарын, көлемін және шектеулерін сипаттаңыз. Data leakage және class imbalance тәуекелдерін бөлек көрсетіңіз.

Оқытушымен талқылау нәтижесінде семестрлік жобаға енгізілетін бір негізгі және бір резервтік ML тәсілін таңдаңыз. Таңдауды метрикалармен қатар есептеу құны, latency және практикалық интеграция мүмкіндігі бойынша негіздеңіз.

Орындау тәртібі

1. Кибершабуылдарды анықтауда AI/ML қолданудың кемінде төрт бағытын жүйелеу.
2. Семестрлік жобаға сәйкес бір нақты ML міндетін тұжырымдау.
3. Кемінде үш модель/тәсілді салыстыру, олардың біреуін baseline ретінде белгілеу.
4. Dataset, feature, label және preprocessing талаптарын сипаттау.
5. Class imbalance және data leakage тәуекелдерін талдау.
6. Бағалау метрикаларын және threshold таңдау логикасын негіздеу.
7. Модельдердің түсіндірмелілігі мен latency талаптарын салыстыру.
8. Негізгі және резервтік ML тәсілін таңдап, оқытушымен талқылауға арналған қысқа негіздеме дайындау.

Тапсырмалар

1. Кибершабуылдарды анықтауда AI/ML қолданудың кемінде төрт бағытын жүйелеу.
2. Семестрлік жобаға сәйкес бір нақты ML міндетін тұжырымдау.
3. Кемінде үш модель/тәсілді салыстыру, олардың біреуін baseline ретінде белгілеу.
4. Dataset, feature, label және preprocessing талаптарын сипаттау.
5. Class imbalance және data leakage тәуекелдерін талдау.
6. Бағалау метрикаларын және threshold таңдау логикасын негіздеу.
7. Модельдердің түсіндірмелілігі мен latency талаптарын салыстыру.
8. Негізгі және резервтік ML тәсілін таңдап, оқытушымен талқылауға арналған қысқа негіздеме дайындау.

Орындау мысалы

Мысалы, желілік ағындарды normal/attack деп жіктеу үшін Logistic Regression baseline ретінде, Random Forest және Gradient Boosting балама ретінде қарастырылады. Бірдей split және preprocessing жағдайында Precision, Recall, F1-score және inference latency салыстырылып, семестрлік жобаға лайық модель таңдалады.

Жеке тапсырма

Әр магистрант өз семестрлік жобасының деректеріне сәйкес AI/ML қолданудың жеке сценарийін ұсынады. Міндет, dataset, модельдер жиыны немесе feature тобы басқа магистранттардың жұмысынан ерекшеленуі тиіс.

Нәтижелерді ұсыну формасы

– 5–7 бет аналитикалық есеп
– AI/ML қолдану бағыттарының картасы
– үш тәсілді салыстыру кестесі
– dataset және feature сипаттамасы
– негізгі/резервтік модель таңдау негіздемесі
– оқытушымен талқылаудан кейінгі қысқа түзету қорытындысы

Есепке қойылатын талаптар

Есепте жұмыстың мақсаты, зерттеу сұрағы, пайдаланылған деректердің немесе қауіпсіздік артефактілерінің шығу тегі, талдау әдісі, салыстыру критерийлері, нәтижелер және олардың интерпретациясы берілуі тиіс. Қорытынды әрбір сандық немесе сапалық нәтижемен байланыстырылып, жұмыстың шектеулері көрсетіледі.

Код қолданылған жағдайда random_state немесе seed, split, preprocessing кезеңдері, негізгі параметрлер және бағдарламалық орта тіркеледі. Нәтижені тек кесте немесе метрика түрінде беру жеткіліксіз: магистрант алынған айырмашылықтардың себебін және нәтижені қандай жағдайда жалпылауға болмайтынын түсіндіреді.

Бағалау критерийлері

Критерий	**Үлес салмағы**
Теориялық негіз және терминдерді дұрыс қолдану	15%
Әдіс пен критерийлерді негіздеу	15%
Деректер/артефактілермен жұмыс	20%
Салыстыру және нәтижелерді талдау	20%
Жеке тапсырма және семестрлік жобамен байланыс	20%
Есепті рәсімдеу және дереккөздер	10%

Бақылау сұрақтары

1. AI/ML кибершабуылдарды анықтауда қандай міндеттерді шеше алады?
2. Supervised және unsupervised тәсілдердің айырмашылығы қандай?
3. Baseline не үшін қажет?
4. Dataset сапасы модельге қалай әсер етеді?
5. Data leakage деген не?
6. Class imbalance қандай мәселе туғызады?
7. Threshold нені өзгертеді?
8. Неліктен F1-score жалғыз өзі жеткіліксіз?
9. Түсіндірмелілік не үшін маңызды?
10. Модельді семестрлік жобаға енгізу кезінде қандай инженерлік критерийлер ескеріледі?

Әдебиеттер тізімі

1. Менщиков А.А., Заколдаев Д.А., Воробьева А.А. Введение в кибербезопасность и киберустойчивость: учебно-методическое пособие. Санкт-Петербург: Университет ИТМО, 2024. 72 с.

2. Бегаев А.Н., Бегаев С.Н., Федотов В.А. Тестирование на проникновение: учебно-методическое пособие. Санкт-Петербург: Университет ИТМО, 2018. 45 с.

3. Вострецова Е.В. Основы информационной безопасности: учебное пособие для студентов вузов. Екатеринбург: Издательство Уральского университета, 2019. 204 с.

4. Шаньгин В.Ф. Информационная безопасность: учебное пособие. Москва: ДМК Пресс, 2014. 702 с.

5. Назарян А.К., Карцан И.Н. Современные кибератаки: классификация и способы защиты // Информатика. Экономика. Управление. 2025. Т. 4, № 1. С. 1001–1007.

6. Лукьянец С.Е. Гибридные системы кибербезопасности: искусственный интеллект и анализ сетевого трафика для противодействия современным угрозам.

7. Сигалов Д.А. Методы выявления поверхности атаки веб-приложений при помощи анализа клиентского JavaScript-кода: диссертация. Москва: МГУ имени М.В. Ломоносова, 2024.

8. Кривчун Н.А., Солодовников Д.В., Соколов А.К., Шорохов К.Д., Кучук М.И. Кибератаки и методы их предотвращения в информационных системах // Дискуссия. 2024. Вып. 128. С. 120–126.