ҒТАМР 81.93.29 **https://doi.org/10.58805/kazutb.v.1.26-743**

ЖЕЛІЛЕРДІҢ ОСАЛДЫҚТАРЫН АНЫҚТАУ ҮШІН КӨП ФАКТОРЛЫ БОЛЖАУ АЛГОРИТМІН БЕЙІМДЕУ

[1]А К. Шегетаева✉, [1]А.Б. Оспанова, [2]Н.С. Смакова, [3]Б. Рысбекқызы, [2]С.А. Алтынбек, [4]Л.Н. Кулбаева [5]Г.Ж. Шуйтенов

[1]Л.Н. Гумилев атындағы Еуразия Ұлттық университеті, Астана, Қазақстан,
[2]Қ.Құлажанов атындағы Қазақ технология және бизнес университеті, Астана, Қазақстан,
[3]Әбілқас Сағынов атындағы Қарағанды техникалық университеті, Қарағанды, Қазақстан,
[4]Astana IT University, Астана, Қазақстан
[5]Esil University, Астана, Қазақстан

✉*Корреспондент-автор: aizhanshegetaeva@mail.ru*

Желілік қауіпсіздік қазіргі цифрлық әлемдегі ең жылдам өзгеретін салалардың бірі болып қала береді. Бұл әсіресе жаңа технологиялардың дамуына және кибершабуылдар санының артуына байланысты өзекті болып табылады. Уақыт өте келе қорғау әдістері мен технологиялары ескіреді және жаңартуды қажет етеді. Бүгінгі таңда өнеркәсіптік желілер мен компьютерлік қауіпсіздік кибершабуылдар санының өсуіне және желінің осалдықтарына байланысты маңыздырақ болып отыр. Сондықтан жыл сайын анықталатын желілік осалдықтарды болжау және автоматтандыру өте маңызды. Бұл зерттеудің негізгі мақсаты - көп факторлы болжау алгоритмі арқылы желідегі осалдықтарды және кибершабуылдарды кешенді зерттеу. Көп факторлы болжау алгоритмі желі осалдықтарының жиілігін, желі трафигінің белсенділігін және қауіпке жауап беру уақытын ескере отырып, шабуылды болжау дәлдігін арттыруға мүмкіндік береді. Зерттеу барысында, Scopus және Web of Science сайттарында жарияланған зерттеу мақалаларына әдебиеттік шолу жүргізу және нейрондық желілер, логистикалық регрессия және кездейсоқ ормандар сияқты машиналық оқыту әдістерін қолдана отырып, деректерді қалыпқа келтіру және талдау арқылы әртүрлі көздерден деректерді жинауды қамтыды. Зерттеудің практикалық маңыздылығы әзірленген алгоритмді желі осалдықтарын бақылау жүйелерін құру үшін пайдалануға болады. Бұл корпоративтік желілердің қауіпсіздігін айтарлықтай жақсартады, кибершабуылдардан болатын зиянды азайтады және жалпы киберқауіпсіздікті жақсартады.

Түйін сөздер: киберқауіпсіздік, көп факторлы болжау, машиналық оқыту, нейрондық желілер , желі осалдықтар, эксплойттар

АДАПТАЦИЯ АЛГОРИТМА МНОГОФАКТОРНОГО ПРОГНОЗИРОВАНИЯ ДЛЯ ОБНАРУЖЕНИЯ СЕТЕВЫХ УЯЗВИМОСТЕЙ

[1]А К. Шегетаева✉, [1]А.Б. Оспанова, [2]Н.С. Смакова, [3]Б. Рысбекқызы, [2]С.А. Алтынбек, [4]Л.Н. Кулбаева [5]Г.Ж. Шуйтенов

[1]Евразийский национальный университет им. Л.Н.Гумилева, Астана, Казахстан,
[2]Казахский университет технологий и бизнеса имени К.Кулажанова, Астана, Казахстан,
[3]Карагандинский технический университет имени Абылкас Сагинова, Караганда, Казахстан,
[4]Astana IT University, Астана, Казахстан,
[5]Esil University, Астана, Казахстан,
e-mail: aizhanshegetaeva@mail.ru

Сетевая безопасность по понятным причинам остается одной из самых быстро меняющихся областей в цифровом мире. Кроме того, в связи с развитием новых технологий и появлением новых кибератак устаревание методов и технологий защиты происходит быстрыми темпами. В настоящее время сетевая и компьютерная безопасность в производственных средах становится все более важной из-за растущего числа кибератак и сетевых уязвимостей. Из-за увеличения количества удаленно подключенных устройств, увеличения объема данных и сложности сетевых технологий. Учитывая большое

количество уязвимостей, выявляемых каждый год, автоматизация их прогнозирования имеет решающее значение. Целью данного исследования является комплексный анализ уязвимостей и атак в сети с использованием многофакторного алгоритма прогнозирования. Алгоритм многофакторного прогнозирования повышает точность прогнозирования атак за счет учета таких параметров, как частота сбоев в работе сети, активность сетевого трафика и время реакции на угрозы. Исследование включало проведение обзора литературы статей, опубликованных в Scopus и Web of Science, сбор данных из различных источников, их нормализацию и анализ с использованием методов машинного обучения, таких как нейронные сети, логистическая регрессия и случайные леса. Практическая ценность исследования заключается в том, что разработанный алгоритм может быть использован для создания систем мониторинга уязвимостей сетей и оценки эффективности эксплуатации в режиме реального времени. Это значительно улучшает защиту корпоративных и государственных сетей, снижает ущерб от кибератак и повышает общую кибербезопасность.

Ключевые слова: многофакторное прогнозирование, машинное обучение, нейронные сети, уязвимости сетей, эксплойты

ADAPTATION OF MULTI-FACTOR FORECASTING ALGORITHM FOR DETECTING NETWORK VULNERABILITIES

[1]A. Shegetayeva✉, [1]A. Ospanova, [2]N. Smakova, [3]B. Rysbekkyzy, [2]S. Altynbek, [4]L. Kulbayeva [5]G. Shuytenov

[1] L.N. Gumilyov Eurasian National University, Astana, Kazakhstan,
[2] K.Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan,
[3]Abylkas Saginov Karaganda Technical University, Karaganda, Kazakhstan,
[4]Astana IT University, Astana, Kazakhstan,
[5]Esil University, Astana, Kazakhstan,
e-mail: aizhanshegetaeva@mail.ru

Network security understandably remains one of the fastest changing areas in the digital world. In addition, with the development of new technologies and the emergence of new cyber attacks, the obsolescence of security methods and technologies occurs at a rapid pace. Nowadays, network and computer security in production environments is becoming increasingly important due to the growing number of cyber attacks and network vulnerabilities. Due to the increase in the number of remotely connected devices, the increase in the volume of data and the complexity of network technologies. Given the large number of vulnerabilities identified each year, automating their prediction is crucial. The objective of this study is to comprehensively analyze network vulnerabilities and attacks using a multi-factor forecasting algorithm. The multi-factor forecasting algorithm improves the accuracy of attack forecasting by taking into account parameters such as network failure rate, network traffic activity and threat response time. The study included a literature review of articles published in Scopus and Web of Science, data collection from various sources, their normalization and analysis using machine learning methods such as neural networks, logistic regression and random forests. The practical value of the study is that the developed algorithm can be used to create systems for monitoring network vulnerabilities and assessing the effectiveness of operation in real time. This significantly improves the protection of corporate and government networks, reduces damage from cyberattacks and improves overall cybersecurity.

Keywords: multivariate forecasting, machine learning, neural networks, network vulnerabilities, exploits

Кіріспе. Соңғы жылдары кибершабуылдар мен желінің осалдығы күрделі жаһандық проблемаға айналды. Қосылған құрылғылардың санының өсуі, деректер көлемінің өсуі және желілік технологиялардың күрделілігі бұл мәселені айқын көрсетеді. Көптеген компаниялар мен ұйымдар қашықтан жұмыс істейтіндіктен, қылмыскерлердің жаңа мүмкіндіктері бар. Мысалы, фи-

шинг, DNS жазбаларын ұрлау, DDoS шабуылдары және құпия сөзді бұзу сияқты қауіптер бүгінгі күні өте өзекті. Сонымен қатар, бағдарламалық қамтамасыз ету қателері, желілік хаттаманың осалдықтары және маңызды жүйелердегі қауіпсіздік мәселелері мен интернеттегі пайдаланушылардың абайсыз әрекеті желінің осалдықтары айтарлықтай қауіп төндіреді.

Кибершабуылдармен күресу әдістері, олардың өсуіне қарамастан айтарлықтай өзгерді. Автоматтандырылған шешімдерді және көп факторлы болжау алгоритмдерін пайдалану желінің осалдықтар мен шабуылдардың алдын алудың маңызды шаралары болып есептелінеді [1]. Интрузияны анықтау жүйелері (IDS), желіаралық қалқандар және антивирустық бағдарламалық қамтамасыз ету көп деңгейлі қорғаныстың негізін құрайды. Бұл шаралар деректерге рұқсатсыз қол жеткізуді болдырмайды, сонымен қатар фишинг және DDoS сияқты шабуылдардан болатын зиянды азайтады [2].

Брандмауэр және шабуылды анықтау жүйелері (IDS) сияқты киберқауіпсіздік жүйелері заманауи шабуылдардың алдын алу үшін жеткілікті тиімді емес. Мысалы, эксплойттар қазір күрделі «камуфляж» әдістерін жиі пайдаланады, бұл анықтау мен бейтараптандыруды қиындатады. Нәтижесінде, қазіргі заманғы машиналық оқыту әдістерін қолданатын және бірнеше факторларды ескеретін кибершабуылдарды болжау және алдын алудың жаңа, дәлірек әдістерін жасау қажеттілігі туындады.

Зерттеудің мақсаты – көп факторлы болжау алгоритмін желінің осалдықтары мен эксплойттарына бейімдеу. Бұл кибершабуылдарды жақсы түсіну және алдын алу, сондай-ақ олар келтіруі мүмкін зияндардан қорғау және азайту стратегияларын әзірлеу үшін қажет. Зерттеудің негізгі міндеттері: 1) жүйелердің осалдығына әсер ететін негізгі факторларды және олардың эксплуатацияларын анықтау; 2) Желілік трафик, шабуыл жиілігі және қауіптерге жауап беру уақыты туралы деректер негізінде көп факторлы шабуылды болжау алгоритмін құру; және 3) желілік трафик деректеріне негізделген алгоритмді әзірлеу.

Материалдар мен әдістер. Заманауи осалдықты бағалау әдістері киберқауіпсіздік тәуекелдерін талдау және басқару үшін әртүрлі әдістер мен алгоритмдерді қолдайды. Графикалық шабуылдарды желілер мен жүйелердегі ықтимал шабуыл жолдарын модельдеу үшін пайдалануға болады. Карталар әлсіз жақтарды анықтауға және шабуылдарды болжауға көмектеседі. Мысалы, графикалық өндіру әдістері IoT желілеріндегі осалдықтарды бағалау үшін Марковтың шешім қабылдау процестері және машиналық оқыту алгоритмдері сияқты модельдерді пайдаланады [3].

Қауіпсіздік мәселелері архитектурада, жобалауда, кодта немесе бағдарламалық қамтамасыз етуді енгізуде пайда болады, жалпы қауіпсіздік ақауларының ресми тізілімі немесе сөздігі бар. Бұл тізім бағдарламалық қамтамасыз етудің қауіпсіздік ақауларын сипаттайтын, сондай-ақ осы ақауларды тануға, жоюға және болдырмауға арналған әмбебап ресми тіл ретінде әрекет етеді. CWE кибершабуылдар тізімі мен осалдықтар деректер қорына NDV, CVE, VulnDB, X-Force және т.б. қоса, осалдықтарды жіктеу мен бағалаудың әртүрлі жүйелері NIPC, SANC, nCirle, CVSS кеңінен қолданылады.

CVSS (Жалпы осалдықты бағалау жүйесі) – осалдықтарды пайдалану ықтималдығы мен желіге әсер ету негізінде бағалаудың жалпы қабылданған классификациясы. Соңғы зерттеулер әрбір инфрақұрылымның сипаттамаларын ескере отырып, нақты желілік орталарда тәуекелді дәлірек бағалау үшін жетілдірілген алгоритмдерді қолдануды көрсетеді [4].

Осалдықтар – жүйенің әлсіз жері, шабуылдаушылар оның қауіпсіздігі мен тұтастығын бұзу үшін пайдалана алады. Осалдықты талдау жүйелердегі ақпараттық қауіпсіздікті қамтамасыз етудің маңызды құралы болып табылады. Ол рұқсат етілмеген қол жеткізу, ақпаратты өзгерту немесе жою үшін шабуылдаушылар пайдаланатын деректер қауіпсіздігін бұзудың ықтимал осалдықтары мен тәуекелдерін анықтауға мүмкіндік береді [5].

Бұл құнды деректер мен активтердің тікелей ағып кетуі (әртүрлі ішкі жүйелерді басқаруға қашықтан қол жеткізу, шоттарды бұзу), соның

ішінде бизнес-процестердің үзілуі, компанияның жұмысын тоқтату (қажет етпейтін шабуылдар) түріндегі елеулі қаржылық шығындар болуы мүмкін. желіге ену де жиі кездеседі (DoS, DDoS). Сондай-ақ, компанияның беделі мен тұтынушылардың сеніміне нұқсан келтіретін бопсалау немесе жариялаумен жалғасатын корпоративтік және құпия ақпараттың (деректер базасын бұзу) шығуы экономикалық шығындарға әкеледі. Жаңа желілік және бұлттық технологиялар, корпоративтік желілердің әртүрлі деңгейлеріндегі инфрақұрылым сөзсіз дерлік белгілі бір осалдықтарға ие, олардың көпшілігін жойылғанға дейін жеке хакерлер немесе тұтас топтар әдейі пайдаланады [6].

Эксплойттар SQL инъекциясы, буфердің толып кетуі, қол жеткізуді басқару осалдықтары және сайтаралық сценарийлер (XSS) осалдықтары сияқты осалдықтардың әртүрлі түрлеріне қарай жіктеледі. Эксплойттар прокси серверлер және қауіпсіздік механизмдерін айналып өту сияқты күрделі әдістерін пайдаланады, бұл шабуылды анықтау мен ізді қалпына келтіруді қиындатады.

Нәтижелер мен талқылау. Болжаудың дәлдігі қазіргі заманғы машиналық оқыту әдістері мен тарихи деректерді пайдалану арқылы артады. Бұл жүйелерді тиімдірек қорғауға және шабуылдардың алдын алуға көмектеседі.

Киберқауіпсіздікте көп факторлы болжау алгоритмі қауіпті болжау мен тәуекелдерді басқарудың қуатты құралы болып табылады. Болжаудың дәлдігі қазіргі заманғы машиналық оқыту әдістері мен тарихи деректерді пайдалану арқылы артады. Бұл жүйелерді тиімдірек қорғауға және шабуылдардың алдын алуға көмектеседі. Көп факторлы болжау алгоритмін құрудың блок-схемасы 1-суретте көрсетілген. Бұл қадамдар деректерді жинауды, деректерді тазалауды және нормалауды, корреляциялық талдауды және факторларды таңдауды, модельді және алгоритмді таңдауды, оқыту мен тестілеуді, болжау және қорытынды жасауды қамтиды.

1 - сурет. Көп факторлы болжауды дамытудың блок-схемасы

Көп факторлы болжау алгоритмін жасау математикалық әдістерді қолдануды және деректерді дұрыс бөлуді талап етеді, толығырақ процесс төменде берілген. Ол деректерді тазалау және қалыпқа келтіру формулаларын, корреляциялық талдауды және үлгілерді оқыту мен сынау әдістерін қамтиды.

1 . Дереккөздерден мәліметтер жинау

Кез келген болжау алгоритмінің негізі - деректер болып табылады. Бұл оқиғалар журналдары, кибершабуылдар мен шабуылдарды анықтау жүйесінің (IDS) дереккорлары және сыртқы осалдық репозиторийлері сияқты әртүрлі көздерден алынуы мүмкін. EXPLOIT - DB (https://www.exploit-db.com/). Зерттеу үшін Exploit-DB дереккорынан алынған 500 желінің осал тұстары мен қатысты эксплойттар үлгісі құрастырылды. Үлгі буфердің толып кетуі, веббағдарламаның осалдықтары (SQL инъекциясы,

XSS), аутентификация және артықшылық қателері, сондай-ақ желі құрылғысы мен қызмет конфигурацияларындағы әртүрлі типтегі осалдықтарды қамтиды. Әрбір эксплуатация бірнеше критерийлер негізінде бағаланады, соның ішінде пайдалану қиындығы, сәттілік ықтималдығы және ықтимал зиян. Деректерді келесі критерийлер бойынша таңдалды, олардың келесі айнымалылары бар: X_1– осалдықтардың жиілігі (белгілі бір кезеңде жүйеде анықталған осалдықтардың саны); X_2– желілік трафик белсенділігі (кіріс және шығыс деректер пакеттерінің көлемі); X_3– шабуыл жиілігі (желіге жазылған шабуылдар саны); X_4– қауіптерге жауап беру уақыты (қауіптерді анықтау және жоюдың орташа уақыты); X_5 – осалдықтардың түрі (CVSS стандартына сәйкес осалдықтардың жіктелуі).

2. Деректерді тазалау және қалыпқа келтіру

Деректерді тазалау көшірмелерді жоюды, жетіспейтін мәндерді жоюды және қателерді түзетуді қамтиды. Аномалияларды жою және деректерді қалыпқа келтіру үшін келесі әдістер қолданылады.

Аномалияларды жою формуласы

$$X_{new} = \frac{X - \mu}{\sigma} \qquad (1)$$

мұндағы X - айнымалының бастапқы мәні, μ - орташа мәні σ және стандартты ауытқуы. (1) формула шектен шығуды жоюға көмектеседі және қалыпты таралуға әкеледі.

Нормалау:

$$X_{norm} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \qquad (2)$$

мұндағы $X_{\min}$ және $X_{\max}$ әрбір айнымалы үшін ең төменгі және ең үлкен мәндер. Бұл процесс деректерді 0-ден 1-ге дейінгі диапазонға қалыпқа келтіріп, машиналық оқыту үлгілерінің жұмысын жеңілдетеді.

3. Корреляциялық талдау және факторларды таңдау

Корреляциялық талдау әртүрлі айнымалылар мен мақсатты айнымалылар арасындағы байланысты анықтауға көмектеседі (мысалы, шабуылдың ықтималдығы). Корреляцияны өлшеудің бір жолы Пирсон коэффициентін қолдану болып табылады:

Пирсон корреляция коэффициентінің формуласы

$$r_{xy} = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2 \sum (Y - \overline{Y})^2}} \qquad (3)$$

мұндағы Y тәуелсіз және X тәуелді айнымалылар, $\overline{X}$ және $\overline{Y}$ олардың орташа мәндері.

Бұл коэффициент факторлар мен нәтиже арасындағы сызықтық байланыстың дәрежесін көрсетеді.

4. Алгоритм мен модельді таңдау

Киберқауіпсіздікте көп факторлы болжау үшін әртүрлі алгоритмдерді қолдануға болады, олардың арасында:

4.1 *Логистикалық регрессия* оқиғаның ықтималдығы болжанатын екілік жіктеу үшін қолданылады (мысалы, шабуыл/шабуылсыз):

$$P(Y = 1|X) = \frac{1}{1 + e^{-z}} \qquad (4)$$

мұндағы $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n$, а $\beta_0, \beta_1, ... \beta_n$- үлгілік оқыту нәтижесінде алынған коэффициенттер.

$P(Y = 1|X)$– оқиғаның орын алу ықтималдығы (мысалы, шабуылдың ықтималдығы).

$X_1, X_2, ..., X_n$– тәуелсіз айнымалылар (факторлар).

$\beta_0, \beta_1, ..., \beta_n$– оқытылатын модельдің параметрлері.

e – көрсеткіш.

4.2 *Шешім ағаштары (Decision Trees)*

Шешім ағаштары шешімдер қабылданатын шарттар (тармақтар) тізбегін құрайды. Тармақ айнымалылардағы шарттарды білдіреді, ал жапырақтарда сынып белгілері бар (мысалы, шабуыл). Алгоритм деректерді энтропия немесе Дженни индексі сияқты критерий негізінде бөледі.

Энтропия:

$$H(p) = -\sum_{i=1}^{C} p_i \log_2(p_i) \quad (5)$$

мұндағы p_i сыныпқа жататын объектілердің үлесі, i және C сыныптар саны.

Джини индексі:

$$G(p) = 1 - \sum_{i=1}^{C} p_i^2 \quad (6)$$

Алгоритм бөлу үшін ең тиімді сипатты таңдау үшін осы көрсеткіштердің бірін азайтуға тырысады.

4.3 *Кездейсоқ орман (Random Forest).*

Кездейсоқ орман - дәлдікті жақсарту және артық орнатуды болдырмау үшін бірнеше шешім ағаштарын біріктіретін ансамбль әдісі. Алгоритм деректердің әртүрлі жиынтықтары бойынша көптеген шешім ағаштарын құрастырады және дауыс беру арқылы соңғы шешімді шығарады:

$$P(Y = 1) = \frac{1}{T} \sum_{t=1}^{T} P_t(Y = 1) \quad (7)$$

мұндағы T ағаштар саны, $P_t(Y = 1)$ сынып үшін ағаштың $Y = 1$ болжаған ықтималдығы t.

4.4 *Нейрондық желілер (Neural Networks).*

Нейрондық желілер қабаттарға біріктірілген есептеу түйіндерін (нейрондарды) көрсету арқылы мидың жұмысын модельдейді. Ең көп тараған түрі – көпқабатты перцептрондар (MLP).

Бір нейронның формуласы:

$$a_j = \sigma\left(\sum_{i=1}^{n} w_{ji} x_i + b_j\right) \quad (8)$$

мұндағы a_j – нейронның шығысы, w_{ji} – i кіріс пен j нейрон арасындағы байланыстың салмағы, x_i – кіріс деректері, b_j – ығысу, σ – белсендіру функциясы (мысалы, сигма тәрізді немесе ReLU).

Сигмоидты белсендіру функциясы:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (9)$$

ReLU (түзетілген сызықтық бірлік)

$$\sigma(z) = \max(0, z) \quad (10)$$

Нейрондық желілер қателіктердің таралуына кері әсер ететін логистикалық жоғалтуды минимизациялайды (мысалы, орташа квадраттық қателіктер).

Бұл алгоритмдер көбінесе шабуыл және осалдықты болжау сияқты киберқауіпсіздік тапсырмалар үшін пайдаланылады.

5. Оқыту және тестілеу

Деректер оқу және сынақ үлгілеріне бөлінуі, әдеттегі деректер 70/30 қатынасын құрайды, мұнда 70% оқыту үшін және 30% тестілеу үшін пайдаланылады. Бұл модельге деректердің бір бөлігінен «үйренуге» және шамадан тыс өзін жаңа ақпаратта сынауға мүмкіндік береді.

Модельді үйрету үшін бинарлық есептеулер үшін логистикалық жоғалту функциясы пайдаланылады:

$$L = \frac{-1}{N} \sum_{i=1}^{N} \left[y_i \log\left(\hat{y}_i\right) + (1 - y_i) \log(1 - \hat{y}_i) \right]$$

мұндағы y_i ақиқат мән, $\hat{y}_i$ болжамды мән және N бақылаулар саны.

Дәлдік, толықтық және F1 өлшемі, классификация моделінің өнімділігін бағалауға арналған негізгі көрсеткіштер болып табылады, әсіресе шабуылдар мен осалдықтарды дұрыс анықтау келесі формулалар арқылы есептеледі:

Дәлдік (Precision):

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

мұндағы TP (True Positive) - дұрыс болжанған оң мысалдар саны (мысалы, анықталған шабуылдар); FP (False Positives) – модель теріс мысалды оң деп қате жіктеген кездегі жалған оң мәндердің саны (мысалы, «жалған дабыл»).

Толықтық (Accuracy):

$$Accuracy = \frac{TP}{TP + FN}$$

мұндағы FN (False Negatives) – модель оң мысалды анықтай алмаған мысалдар саны (мысалы, өткізіп алған шабуылдар саны).

F1-өлшем (F1-балл):

$$F1 - score = 2\frac{Precision \times Accuracy}{Precision + Accuracy} \quad (12)$$

Жоғарыдағы (12)-ші формулада F1-өлшемі дәлдік пен толықтық арасындағы гармоникалық орташа мән болып табылады, ол модельді неғұрлым теңгерімді бағалау үшін қолданылады, әсіресе жалған позитивтерді де, өткізіп алған шабуылдарды да есепке алу маңызды.

6. Болжау және қорытындылар

Модель оқытылып, тексерілгеннен кейін, оны болжау үшін пайдалануға болады. Болжау кіріс ретінде қабылдайтын және оқиғалардың ықтималдығын шығаратын болжау функциясына негізделген маңызды фактор.

500 желі осалдықтарының үлгісі мен Exploit-DB дерекқорынан эксплойттардың тиімділігін X_1- X_5 көрсеткіштерге сәйкес сипаттау үшін 1 - кестені әрбір индикатор үшін максималды, ең төменгі және орташа мәндермен ұсынылды.

1 - кесте. Зерттелетін көрсеткіштердің максималды, ең төменгі және орташа мәндері

Көрсеткіш	Өлшем бірліктері	Максималды мән	Ең төменгі мән	Орташа мән (η)
X_1	тәулігіне жазбалар саны (бірлік/күн)	200	1	50
X_2	желілік трафик көлемі (ГБ/күн)	800	10	250
X_3	күніне шабуылдар саны (бірлік/күн)	500	20	150
X_4	қауіптерге жауап беру уақыты (минут)	600	5	180
X_5	CVSS осалдық көрсеткіші (баллдар)	10	3	6.5

2 - кесте. Деректерді тазалау (аномалияларды жою)

Көрсеткіш	Орташа (η)	Стандартты ауытқу (σ)	X_{new} (аномалияларды жойғаннан кейін)
X_1	50	70	(X-50) / 70
X_2	250	200	(X-250) / 200
X_3	150	120	(X-150) / 120
X_4	180	160	(X-180) / 160
X_5	6.5	2	(X-6,5) / 2

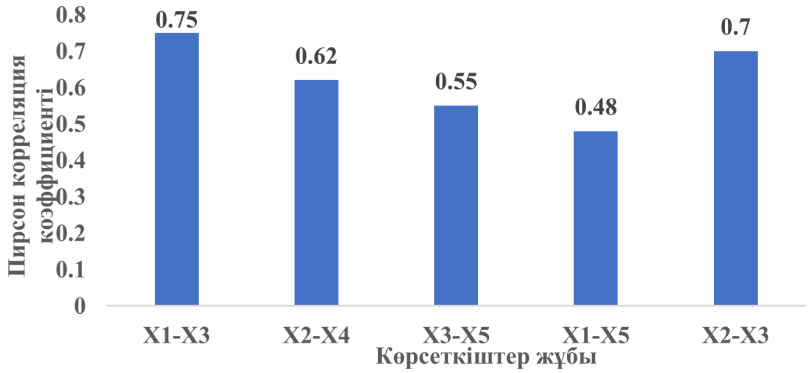
2 - сурет. 1 тәулікте болатын осалдықтар мен шабуылдардың максималды және минималды көрсеткіштері

1-ші кесте жаңа өлшем бірліктерін ескере отырып, осалдықтардың пайда болу жиілігін, шабуылдар жиілігін және қауіптерге жауап беру уақытын түсіндірудегі өзгерістерді көрсетеді (2 сурет). 2-ші кесте деректерді тазалау (аномалияларды жою) мысалын көрсетеді.

Деректерді қалыпқа келтіру үшін (2) формула пайдаланылды. Деректерді қалыпқа келтіру нәтижелері 3-ші кестеде көрсетіледі.

3 - кесте. Деректерді қалыпқа келтіру

Көрсеткіш	Ең аз (X_{min})	Максимум(X_{max})	Шамамен нормаланған мән (X_{norm})
X_1	1	200	(X-1) / (200-1)
X_2	10	800	(X-10) / (800-10)
X_3	20	500	(X-20) / (500-20)
X_4	5	600	(X-5) / (600-5)
X_5	3	10	(X-3) / (10-3)



3 - сурет. Пирсон корреляция коэффициенттері

X_1-X_5 көрсеткіштер және мақсатты айнымалылар арасындағы корреляцияны талдау жүргізу үшін талдау Пирсон корреляция коэффициенті (3) формуласы пайдаланылды. Корреляциялық

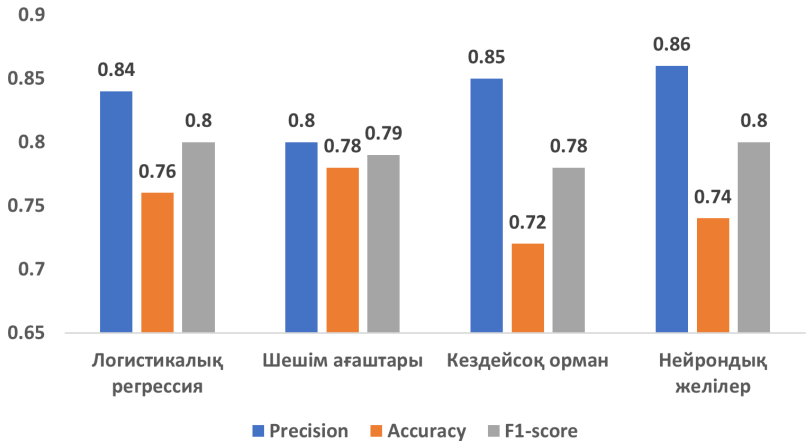
талдаудың нәтижелері 3-ші суретте көрсетіледі.

Ең жоғары оң корреляция ($r_{xy} = 0.75$) – X_1 (осалдықтардың пайда болу жиілігі) және X_3 (шабуылдар жиілігі) арасында байқалады, бұл осалдықтар саны мен шабуылдар саны арасындағы өзара байланысты көрсетеді.

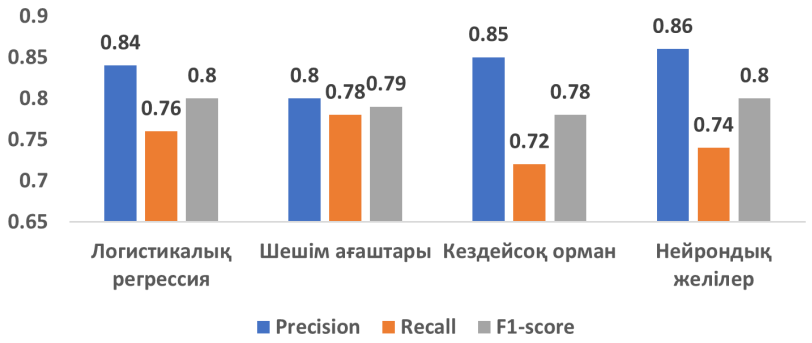
Орташа корреляция ($r_{xy} = 0.62$) – X_2 (желі трафик белсенділігі) және X_4 (қауіпке жауап беру уақыты) арасындағы трафик көлемі мен қауіпке жауап беру уақыты арасындағы тәуелділікті көрсете алады.

Төмендеген корреляция $r_{xy} = 0.55 - 0.48$ – X_3 (күніне шабуылдар саны) және X_5 (CVSS бойынша осалдықты бағалау) көрсеткіштері арасында және X_1 (осалдықтардың жиілігі) және X_5 (CVSS бойынша осалдық көрсеткіші) индикаторлар арасында төмендеген корреляция бар екенін де атап өтуге болады. X_2 (желінің трафик белсенділігі) және X_3 (шабуыл жиілігі) арасындағы корреляция коэффициенті $r_{xy} = 0.70$ жеткілікті жоғары екенін белгілеп, айта кеткен жөн. Бұл нәтижелер көп факторлы болжау үшін негізгі факторларды таңдауға мүмкіндік береді.

3 және 4-суреттер шабуылдың ықтималдығын болжау үшін төрт машиналық оқыту алгоритмін пайдалану нәтижелерін көрсетеді. Дәлдік (Precision), толықтық (Accuracy) және F1-өлшемі (F1-балл) көрсеткіштері оқыту және сынақ үлгілері үшін есептелінді.



4-сурет. Зерттеу үлгісінде алгоритмдерін қолдану нәтижелері



5 - сурет. Сынақ үлгісінде алгоритмдерді қолдану нәтижелері

4 суретте көрсетілген деректерді талдау арқылы, Логистикалық регрессия 0,85 дәлдігін көрсетеді, яғни болжанған шабуылдардың 85% ақиқат мәнінде шабуылдар болды. Толықтығы 0,78 деңгейінде барлық нақты шабуылдардың 78% дұрыс болжағанын көрсетеді. F1-өлшемі 0,81 деңгейі

дәлдік пен толықтық арасындағы жақсы теңгерімді көрсетеді.

Шешім ағаштары дәлдігі 0,82 және толықтық 0,80 сәл теңдестірілген нәтижелерді көрсетеді, бұл 0,81 F1-өлшемінде де көрінеді. Бұл модель шабуылдарды болжау міндетін тиімді екенін білдіреді.

Кездейсоқ орман барлық алгоритмдер арасында ең жоғары 0,88 дәлдікті көрсетті, бұл дұрыс болжамдардың жоғары деңгейін көрсетеді. Алайда, толықтығы 0,74 болды, бұл модель шабуылдарды шамамен 26% өткізіп алғанын көрсетеді. F1-өлшемі 0,80-ге тең, жоғары дәлдікке қарамастан, дәлдік пен толықтық арасындағы тепе-теңдік идеалды емес екенін көрсетеді.

Нейрондық желілер 0,87 дәлдігін көрсетеді, бұл кездейсоқ орман нәтижелеріне жуық, ал еске түсіру 0,76. Бұл нейрондық желілер барлық шабуылдардың 76% дұрыс анықтағанын білдіреді және F1-өлшемі 0,81-ге тең дәлдік пен толықтық арасындағы теңгерімді көрсетеді.

5 суретінде көрсеткендей, Сынақ үлгісінде алгоритмдерді қолдану нәтижелері бойынша Логистикалық регрессия 0,84 дәлдікпен 0,76 толықтық берді, бұл Оқыту үлгісінің нәтижесінен сәл төмен, бірақ F1-өлшемі 0,80 модель жаңа деректерге шабуылдарды болжауда әлі де тиімді.

Сынақ үлгісінде Шешім ағаштары 0,80 дәлдігін және 0,78 толықтықты көрсетті, бұл олардың оқыту үлгісінің нәтижелеріне ұқсас. F1 өлшемі 0,79 мәнді алгоритмнің әртүрлі деректер жиындарында тұрақты орындалатынын растайды, бұл жеткілікті теңдестірілген нәтижелер береді.

Кездейсоқ орман 0,85 дәлдікті көрсетеді, бұл жаттығу жиынындағы нәтижеден небәрі 0,03 төмен. Дегенмен, толықтық 0,72-ге дейін төмендейді, бұл модельдің шабуылдарды анықтау сезімталдығының төмендеуін көрсетеді. F1-өлшемі 0,78-ге тең жаңа деректермен жұмыс істеу кезінде модельдің жалпы балансының төмендеуін көрсетеді.

Нейрондық желілер сынақ үлгісінде 0,86 дәлдікпен тұрақтылықты көрсетті, бұл олардың жаңа деректерге шабуылдарды болжаудың жоғары қабілетін растайды. Толықтығы 0,74-ке дейін төмендеді, бірақ F1-өлшемі 0,80-ге тең, модель әлі де болжамдылықтың жоғары деңгейін және дәлдік пен толықтық арасындағы тепе-теңдікті сақтайтынын растайды.

4 және 5 суреттегі салыстырулар нәтижелерінен, нейрондық желілер шабуылды болжаудың жоғары дәлдігін және теңдестірілген F1-өлшемдерін көрсете отырып, екі үлгіде де ең жақсы нәтижелерді көрсетеді деп қорытынды жасауға болады. Кездейсоқ орман оқыту үлгісінде ең жоғары дәлдікті көрсетті, бірақ оның сынақ үлгісіндегі деректерді жалпылау қабілеті нейрондық желілерге қарағанда төмен болды. Логистикалық регрессия және шешім ағаштары нейрондық желілерге қарағанда тұрақтырақ, бірақ дәлірек нәтижелерді береді.

Машиналық оқытуды пайдаланатын көп факторлы болжау алгоритмі желінің осал тұстарын бағалауда жоғары дәлдікті көрсетті. Бұл осалдықтардың жиілігі, желілік трафик белсенділігі, қауіптерге жауап беру уақыты және басқа да факторлар сияқты әртүрлі параметрлерді ескеру мүмкіндігіне байланысты. Бұл айнымалылар арасындағы күрделі қатынастарды жиі есепке алмайтын дәстүрлі бір факторлы әдістерден оның артықшылығын растайды [7].

Зерттеулер көрсеткендей, осалдықтардың жоғары жиілігі (мысалы, тәулігіне осалдықтардың X_1 саны) жүйелерге жасалған шабуылдардың жиілігімен сәйкес келеді. Сонымен қатар, біздің нәтижелеріміз сәтті шабуылдардың осал желілерді, әсіресе желілік белсенділік жағдайында көбірек нысанаға алатынын көрсетті. Басқа авторлар инфрақұрылымға шабуылдардың жиілігі мен желілік трафик белсенділігі арасындағы байланысты көрсетті, бұл нәтижелерді растайды [8].

Нейрондық желілер және шешім ағаштары сияқты заманауи машиналық оқыту әдістерін пайдалану арқылы көп факторлы болжамды модельдер ықтимал қауіптерді дәлірек анықтай алады. Бұл алгоритмдер үлкен көлемдегі деректерді өңдей алады және дәстүрлі талдау әдістері анықтай алмайтын жасырын үлгілерді таба алады [9].

Көп факторлы болжау алгоритмінің үлкен

артықшылығы, оның жоғары дәлдігі және әртүрлі желі жағдайларына бейімделу мүмкіндігі болып табылады. Дегенмен, оның кемшіліктері бар, оларды одан әрі оңтайландыру үшін ескеру қажет.

Шабуыл жиілігі мен осалдықтар сияқты әртүрлі факторлар қосылғанда, әсіресе нейрондық желілер мен кездейсоқ ормандарды пайдаланғанда, болжау дәлдігі айтарлықтай артады. Біздің экспериментіміз модель болашақ қауіптерді дәл болжай алатынын көрсетті, өйткені сынақ деректері F1-өлшемін 0,80 көрсетті. Алгоритм бірнеше түрлі желілік инфрақұрылымдар мен кибер орталарда жұмыс істейді. Мысалы, бірнеше зерттеулер оның шағын желілерді де, кәсіпорын инфрақұрылымдарын да сәтті модельдейтінін көрсетті. Кибершабуылдарды болжаудағы тиімділігін растай отырып, нейрондық желілер дәлдікте де, F1-өлшемінде де ең жақсы нәтиже көрсетті. Нейрондық желілер күрделі желілік деректерді талдау кезінде жоғары дәлдік көрсетті, бұл осы нәтижелерді растайды [10].

Көп факторлы модельдерді, әсіресе нейрондық желілерге негізделген модельдерді пайдалану үлкен есептеу ресурстарын қажет етеді, бұл оларды нақты уақытта пайдалануды қиындатады. Олардың қаншалықты дәл екендігіне қарамастан, кездейсоқ орман және шешім ағаштары сияқты машиналық оқыту үлгілері, әсіресе шағын деректермен жұмыс істегенде, шамадан тыс орнатуға бейім. Бұл олардың жаттығу деректерінде анықталмаған жаңа шабуылдарды дәл болжау қабілетін төмендетуі мүмкін [11].

Қорытынды. Бұл зерттеудің мақсаты көп факторлы болжау алгоритмі арқылы желінің осалдықтары мен эксплуатацияларының өнімділігін бағалау болды. Осы мақсатқа жету үшін желі белсенділігінің әртүрлі параметрлерін, шабуылдардың жиілігін және қауіпке реакция уақытын ескере отырып, болашақ кибершабуылдарды дәлірек болжауға мүмкіндік беретін алгоритм әзірленді және қолданылды [12].

Зерттеуге арналған деректер оқиғалар журналдары және осалдық дерекқорлары сияқты әртүрлі көздерден жиналды. Деректерді қалыпқа келтіру және тазалаудан кейін логистикалық регрессия, нейрондық желілер, кездейсоқ ормандар және шешім ағаштары сияқты машиналық оқыту әдістері қолданылды. Осалдықтардың түрі, трафик белсенділігі және осалдықтардың жиілігі осы алгоритмдер шабуылдардың ықтималдығын болжау үшін пайдаланатын факторлар болды.

Негізгі зерттеу нәтижелері нейрондық желілер шабуылдарды болжауда тамаша екенін көрсетті. Нейрондық желілер шабуылдарды дәл болжау және сынақ жиынындағы жалған позитивтердің санын азайту қабілетін дәлелдеді, дәлдік 86% және F1-өлшемі 0,80. Екінші жағынан, шешім ағаштары 80% дәлдік пен F1-өлшемі 0,79, ал кездейсоқ ормандар 85% және F1- өлшемі 0,78 дәлдік көрсетті . Бұл осалдықтардың жиілігі мен шабуылдардың жиілігі бір-бірімен тығыз байланысты екенін көрсетті және олардың арасындағы ең жоғары корреляция r = 0,75 болды.

Алынған нәтижелер деректердің ағып кетуін және басқа кибершабуылдарды болдырмау үшін шабуылдарды дәл болжау қажет болатын корпоративтік және мемлекеттік желілерді қорғау саласында қолданылуы мүмкін. Алгоритмді нақты уақыт режимінде желі белсенділігін бақылау жүйелерін құру үшін пайдалануға болады, бұл қауіптерге жауап беру уақытын қысқартуға және инфрақұрылымның жалпы қауіпсіздігін арттыруға мүмкіндік береді.

Әдебиеттер

1. Сасенов А.Б. Прогнозирование развития криминальной ситуации в Республике Казахстан: монография/КариповаА.Т. Касимова М.О., Абулгазина А.Ж., Казбекова А.Б., Тулеуов Б.И., Сыздыков А.Ж., Муканов Д.Ж. Под общей редакцией Ахметзакирова Н.Р. - Астана: Академия правоохранительных органов при Генеральной прокуратуре Республики Казахстан, 2017.- 172 с. ISBN 978-601-06-4184-6

2. Falkevych, V., Lisnyak, A. Internal and External Threats in Cyber Security and Methods for Their

Prevention // In 2023 13th International Conference on Advanced Computer Information Technologies (ACIT). -2023. -P.414 - 419. DOI 10.1109/ACIT58437.2023.10275516

3. Almazrouei, O., Magalingam , P., Hasan, M., & Shanmugam, M. A Review on Attack Graph Analysis for IoT Vulnerability Assessment: Challenges, Open Issues, and Future Directions // IEEE Access. -2023. -Vol.11. -P. 44350-44376. DOI 10.1109/ACCESS.2023.3272053

4. Wang, W., Shi, F., Zhang, M., Xu, C., & Zheng, J. A Vulnerability Risk Assessment Method Based on Heterogeneous Information Network // IEEE Access. -2020. -Vol. 8. -P.148315–148330. DOI 10.1109/ACCESS.2020.3015551

5. Хромова А.Р., Петросян Л.Э. Анализ уязвимостей в системах безопасности данных // Инженерный вестник Дона. -2023. -№ 6.(102)-С.67-76

6. Оспанова А.Б., Шегетаева А.К., Түсіпханов А. Т., Жалгасбаев А. Б., Кадринов Д. М. Прогнозирование сетевых уязвимостей и эксплойтов // Международная научно-практическая конференции «XVI Сагиновские чтения. Интеграция образования, науки и производства». - Караганда: Изд-во КарТУ им. А.Сагинова, 2024. -Ч.2. -С. 287-289.

7. Carriegos, M., Castañeda, Á., Trobajo , M., & Zaballa , D. On Aggregation and Prediction of Cybersecurity Incident Reports //IEEE Access.-2021.-Vol.9.-P.102636-102648. DOI 10.1109/ACCESS.2021.3097834

8. Demidova, L., & Stepanov, M. Development of Multifactor Forecasting Model based on Fuzzy Time Series // 2021 International Conference on Information Technologies (InfoTech). -2021. DOI 10.1109/InfoTech52438.2021.9548389

9. Injadat, M., Moubayed, A., & Shami, A. Detecting Botnet Attacks in IoT Environments: An Optimized Machine Learning Approach // 2020 32nd International Conference on Microelectronics (ICM). -2020. DOI 10.1109/ICM50269.2020.9331794

10. Pavlov, A. Analysis of Network Interaction of Modern Exploits // Informacionnye Tehnologii. -2022. -Vol. 28(2). DOI 10.17587/it.28.75-80

11. Aota, M., Kanehara, H., Kubo, M., Murata, N., Sun, B., Takahashi, T. Automation of Vulnerability Classification from its Description using Machine Learning //2020 IEEE Symposium on Computers and Communications (ISCC). -2020. DOI 10.1109/ISCC50000.2020.9219568

12. Шегетаева А.К. Всесторонний обзор многофакторного прогнозирования сетевых уязвимостей // International Scientific Symposium Karabakh and West Azerbaijan: Triumph of Victory. The 26th of October 2024. - P. 553-560. ISBN 978-625-98125-4-0

References

1. Sasenov A.B. Prognozirovanie razvitija kriminal’ noj situacii v Respublike Kazahstan: monografija / KaripovaA.T. Kasimova M.O., Abulgazina A.Zh., Kazbekova A.B., Tuleuov B.I., Syzdykov A.Zh., Mukanov D.Zh. Pod obshhej redakciej Ahmetzakirova N.R. - Astana: Akademija pravoohranitel’ nyh organov pri General’ noj prokurature Respubliki Kazahstan, 2017.- 172 s. ISBN 978-601-06-4184-6. [in Russian]

2. Falkevych, V., Lisnyak, A. Internal and External Threats in Cyber Security and Methods for Their Prevention // In 2023 13th International Conference on Advanced Computer Information Technologies (ACIT). -2023. -P.414 - 419. DOI 10.1109/ACIT58437.2023.10275516

3. Almazrouei, O., Magalingam , P., Hasan, M., & Shanmugam, M. A Review on Attack Graph Analysis for IoT Vulnerability Assessment: Challenges, Open Issues, and Future Directions // IEEE Access. -2023.

-Vol.11. -P. 44350-44376. DOI 10.1109/ACCESS.2023.3272053

4. Wang, W., Shi, F., Zhang, M., Xu, C., & Zheng, J. A Vulnerability Risk Assessment Method Based on Heterogeneous Information Network // IEEE Access. -2020. -Vol. 8. -P.148315–148330. DOI 10.1109/ACCESS.2020.3015551

5. Hromova A.R., Petrosjan L.Je. Analiz ujazvimostej v sistemah bezopasnosti dannyh // Inzhenernyj vestnik Dona. -2023. -№ 6.(102)-C.67-76. [in Russian

6. Ospanova A.B., Shegetaeva A.K., Tүsiphanov A. T., Zhalgasbaev A. B., Kadrinov D. M. Prognozirovanie setevyh ujazvimostej i jeksplojtov // Mezhdunarodnaja nauchno-prakticheskaja konferencii «XVI Saginovskie chtenija. Integracija obrazovanija, nauki i proizvodstva». - Karaganda: Izd-vo KarTU im. A.Saginova, 2024. -Ch.2. -S. 287-289.[in Russian]

7. Carriegos, M., Castañeda, Á., Trobajo , M., & Zaballa , D. On Aggregation and Prediction of Cybersecurity Incident Reports //IEEE Access.-2021.-Vol.9.-P.102636-102648. DOI 10.1109/ACCESS.2021.3097834

8. Demidova, L., & Stepanov, M. Development of Multifactor Forecasting Model based on Fuzzy Time Series // 2021 International Conference on Information Technologies (InfoTech). -2021. DOI 10.1109/InfoTech52438.2021.9548389

9. Injadat, M., Moubayed, A., & Shami, A. Detecting Botnet Attacks in IoT Environments: An Optimized Machine Learning Approach // 2020 32nd International Conference on Microelectronics (ICM). -2020. DOI 10.1109/ICM50269.2020.9331794

10. Pavlov, A. Analysis of Network Interaction of Modern Exploits // Informacionnye Tehnologii. -2022. -Vol. 28(2). DOI 10.17587/it.28.75-80

11. Aota, M., Kanehara, H., Kubo, M., Murata, N., Sun, B., Takahashi, T. Automation of Vulnerability Classification from its Description using Machine Learning //2020 IEEE Symposium on Computers and Communications (ISCC). -2020.
DOI 10.1109/ISCC50000.2020.9219568

12. Shegetaeva A.K. Vsestoronnij obzor mnogofaktornogo prognozirovanija setevyh ujazvimostej // International Scientific Symposium Karabakh and West Azerbaijan: Triumph of Victory. The 26th of October 2024. - P. 553-560. ISBN 978-625-98125-4-0.[in Russian]

Авторлар туралы мәліметтер

Шегетаева А. К. - докторант, Л.Н. Гумилев атындағы Еуразия Ұлттық университеті, Астана, Қазақстан, e-mail: aizhanshegetaeva@mail.ru;

Оспанова А. Б. - ф. -м.ғ.к., Л.Н. Гумилев атындағы Еуразия Ұлттық университеті, Астана, Қазақстан, e-mail: o.ademi111@gmail.com;

Смакова Н.С.-PhD, ассоц. профессор, Қ.Құлажанов атындағы Қазақ технология және бизнес университеті, Астана, Қазақстан,, Астана, Қазақстан, e-mail: nuri_5@mail.ru;

Рысбекқызы Б. - PhD, аға оқытушысы, Әбілқас Сағынов атындағы Қарағанды техникалық университеті, Қарағанды, Қазақстан, e-mail: Bakhytgulz@mail.ru;

Алтынбек С. А. - PhD, ассоц. профессор, Қ.Құлажанов атындағы Қазақ технология және бизнес университеті, Астана, Қазақстан,, Астана, Қазақстан, e-mail: serik_aa@bk.ru;

Кулбаева Л.Н. - магистр, Astana IT University, Астана, Қазақстан, e-mail: laukakn@mail.ru

Шуйтенов Г.Ж. - доцент кафедры ИСиТ Esil University, Астана, Қазақстан, e-mail: g.shuitenov@mail.ru

Information about authors

Shegetayeva A. - doctoral student, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, e-mail: aizhanshegetaeva@mail.ru;

Ospanova A. - candidate of physical and mathematical sciences, senior lecturer, L.N. Gumilyov Eurasian National University,

Astana, Kazakhstan, e-mail: o.ademi111@gmail.com;

Smakova N. - PhD, K.Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan, e-mail: nuri_5@mail.ru;

Rysbekkyzy B. - PhD, senior lecturer, Abylkas Saginov Karaganda Technical University, Karaganda, Kazakhstan, e-mail: Bakhytgulz@mail.ru;

Altynbek S. - PhD, K.Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan, e-mail: serik_aa@bk.ru;

Кулбаева Л. - магистр, Astana IT University, Astana, Kazakhstan, e-mail: laukakn@mail.ru

Shuytenov G. - associate professor of the department of 'Information Systems and Technologies' at Esil University, Astana, Kazakhstan, e-mail: g.shuitenov@mail.ru