ТЕХНИКАЛЫҚ ҒЫЛЫМДАР
ТЕХНИЧЕСКИЕ НАУКИ
TECHNICAL SCIENCE

CENTRAL ASIAN SCIENTIFIC JOURNAL
ISSN 3135-3061
2026. №4 (32), 2-том. 395–401 б.
DOI: https://doi.org/10.5281/zenodo.24272669
Журнал сайты: https://cajournal.kz

ӘОЖ 004.85

ГИДРАВЛИКАЛЫҚ ЖҮЙЕНІҢ ЖҰМЫС РЕЖИМІНДЕГІ АУЫТҚУЛАРДЫ МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІМЕН АНЫҚТАУ

А.А. Тусупова
Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан

Тусупова Аружан Алмазовна — магистрант, 1 курс, Цифрлық ғылымдар және жасанды интеллект институты, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан.

Ғылыми жетекші: Ералиева Бахыт Шынжырбековна — PhD, доцент м.а, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан.

© А.А. Тусупова, 2026

Аңдатпа: Мақалада гидравликалық сынақ жүйесіндегі жұмыс режимінің ауытқуларын анықтау үшін Isolation Forest, One-Class SVM және Local Outlier Factor әдістері салыстырылады. Деректер ретінде UCI репозиторийіндегі 2205 жұмыс циклінің қысым, температура және шығын бойынша орташа мәндері алынды; модельдер клапан қалыпты күйдегі циклдерде оқытылып, шешім шегі валидация жиынтығында таңдалды. Көлемі 441 циклдік тест жиынтығында Isolation Forest F1-score бойынша ең жақсы нәтиже берді (80,85%, Recall — 92,07%), ал One-Class SVM мен LOF үшін бұл көрсеткіш 50,12% және 58,49% болды. Нәтижелер клапан күйін анықтауға қатысты және сорғының ішкі ағып кетуін диагностикалауға тікелей көшірілмейді.

Түйін сөздер: гидравликалық жүйе, машиналық оқыту, аномалияларды анықтау, Isolation Forest, One-Class SVM, LOF, салыстырмалы талдау

Кіріспе

Гидравликалық жетектер станоктарда, престерде, көтеру-тасымалдау техникасында және технологиялық желілерде жұмыс істейді, ал олардың бұзылуы көбіне күтпеген тоқтап қалумен аяқталады. Жүйенің жағдайы қысымның, сұйық температурасының және шығынның өзгерісінде бірден көрінеді: клапан толық ашылмаса, сорғының ішінен сұйық ағып кетсе немесе салқындату нашарласа, бұл үш шама әдетте бірге өзгереді. Сондықтан датчиктердің көрсеткіштерін үздіксіз талдау жабдықтың жағдайын бағалаудың негізгі құралдарының бірі болып саналады [1].

Өндірісте көбіне қарапайым шекті бақылау қолданылады: әр көрсеткіш алдын ала белгіленген рұқсат етілген аралықпен салыстырылады. Бұл тәсіл түсінікті және есептеуге жеңіл, бірақ бірнеше датчиктің бір мезгілдегі өзгерісін ескермейді: әр шама жеке алғанда нормада тұруы мүмкін, ал олардың үйлесімі әдеттен тыс болады. Машиналық оқыту осындай көпөлшемді ауытқуларды іздеуге мүмкіндік береді [2; 3].

Isolation Forest әдісінің идеясы қарапайым: кездейсоқ таңдалған белгі бойынша кездейсоқ шекпен бөлгенде әдеттен тыс бақылау басқаларға қарағанда тезірек жеке қалады, яғни оны оқшаулау үшін аз бөлу жеткілікті [4]. Әдіске ақау түрлерінің белгіленген мысалдары қажет емес. Алайда мұның екінші жағы бар: аномалия мен нақты техникалық ақау бір нәрсе емес. Жүктеме немесе температура режимінің қалыпты өзгерісін де алгоритм сирек құбылыс ретінде белгілеп қоюы мүмкін.

Осы деректер жиынтығында One-Class SVM, LOF және Isolation Forest әдістері жартылай бақыланатын қойылымда бұрын да салыстырылған [5]. Біз үш әдісті басқаша жағдайда қарастырдық: бірдей белгілер, бірдей бөліну және валидацияда таңдалатын шешім шегі. Модельдер қалыпты клапан күйіндегі деректермен оқытылды, ал нәтижелер ортақ тест жиынтығында бағаланды.

Зерттеудің мақсаты - гидравликалық жүйенің қысым, температура және шығын көрсеткіштері бойынша Isolation Forest, One-Class SVM және LOF алгоритмдерінің тиімділігін салыстыру және шешім шегін таңдаудың нәтижеге әсерін бағалау. Бұл жұмыс болашақта сорғы қондырғысына арналған мониторинг модулін жасауға негіз болады. Мақсатқа жету үшін ашық деректер жиынтығын сипаттап, цикл бойынша белгілерді қалыптастыру, үш алгоритмді бірдей деректерде оқыту, валидацияда шекті таңдау, тест жиынтығында Precision, Recall және F1-score көрсеткіштерін салыстыру және нәтижелердің қолданылу шектерін көрсету қажет болды.

Практикада аномалияны тауып қою ғана жеткіліксіз: жалған ескертулер көп болса, оператор оларға сенуін тоқтатады. Сондықтан сапа Precision, Recall және F1-score бойынша қатар бағаланды, ал шешім шегі валидацияда негізделді. Дегенмен бір деректер жиынтығындағы нәтиже үш әдістің кез келген өндірістік жағдайдағы артықшылығын дәлелдемейді.

Материалдар мен әдістер

Деректер көзі ретінде Helwig, Pignanelli және Schütze жариялаған «Condition Monitoring of Hydraulic Systems» жиынтығы алынды [6]. Деректер гидравликалық сынақ стендінде тіркелген: негізгі жұмыс контуры мен салқындату-сүзгілеу контуры бар жүйеде жүктеме циклдері қайталанып отырады. Әр цикл 60 секундқа созылады, барлығы 2205 цикл тіркелген. Датчиктердің өлшеу жиілігі әртүрлі: PS1 — 100 Гц, FS1 — 10 Гц, TS1 — 1 Гц (1-кесте) [6].

Әр цикл үшін үш сигналдың арифметикалық орташа мәні есептелді, сондықтан бастапқы уақыттық қатарлар 2205 × 3 өлшемді белгілер матрицасына келтірілді. Бұл есептеуді жеңілдетеді, бірақ қысқа мерзімді соққылар мен тербелістер туралы ақпараттың бір бөлігі жоғалады.

Кесте 1 – Зерттеуге таңдалған датчиктер

Белгі	Физикалық шама	Өлшем бірлігі	Жиілік
PS1	Қысым	bar	100 Гц
TS1	Температура	°C	1 Гц
FS1	Көлемдік шығын	л/мин	10 Гц

Мақсатты белгілерге ерекше назар аудару керек. profile.txt файлында бірінші баған салқындатқыштың күйін, екінші баған клапанның күйін (100, 90, 80, 73), үшінші баған сорғының ішкі ағып кетуін (0, 1, 2), төртінші баған гидравликалық аккумулятордың күйін сипаттайды [6]. Бастапқы есептеуде екінші баған қате түрде Pump_condition деп аталған, бірақ сандық есептеулер дәл осы баған бойынша жүргізілген. Мақалада оны «клапан күйі» деп атаймыз; сорғының ішкі ағып кетуі бойынша жеке сапа көрсеткіштері есептелген жоқ.

Есептеулер Python тілінде pandas, NumPy, Matplotlib және scikit-learn [7] кітапханаларымен орындалды. Нәтижені қайталай алу үшін кездейсоқтық тұқымы random_state = 42 деп бекітілді.

Модельдерге беруден бұрын әр белгі стандартталды. Цикл векторы $x_i = (P_i, T_i, Q_i)$ түрінде жазылады, мұндағы P — қысым, T — температура, Q — шығын. Стандарттау формуласы:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

мұндағы μ мен σ — тек оқыту жиынтығында есептелген орташа мән және стандарттық ауытқу; валидация мен тест деректеріне дәл осы параметрлер қолданылды.

Isolation Forest кездейсоқ белгіні және оның мәндер аралығынан кездейсоқ бөлу нүктесін таңдап, оқшаулау ағаштарын салады. Бақылаудың аномалиялық баллы ағаштардағы орташа жол ұзындығына байланысты:

$$s(x, n) = 2^{-\frac{E[h(x)]}{c(n)}} \quad (2)$$

мұндағы $E[h(x)]$ — x бақылауының орташа жол ұзындығы, $c(n)$ — n бақылау үшін нормалаушы коэффициент [4]. scikit-learn іске асыруында score_samples функциясының мәні неғұрлым аз болса, бақылау соғұрлым аномалияға ұқсас болады [7].

Бастапқы тәжірибеде n_estimators = 100 және contamination = 0,05 параметрлерімен модель барлық 2205 циклде оқытылды; оның нәтижесі 1- және 2-суреттерде көрсетілген. Негізгі салыстыру үшін деректер кездейсоқ түрде 60% оқыту, 20% валидация және 20% тест бөліктеріне бөлінді. Оқыту бөлігінен клапаны қалыпты күйдегі (100) циклдер ғана алынды. Валидацияда оқыту жиынтығындағы баллдардың 1, 5, 10, 15, 20, 30, 40 және 50-процентильдері шек ретінде сыналып, F1-score ең жоғары болған шек таңдалды. Тест жиынтығы шекті таңдауға қатысқан жоқ.

Оң класс ретінде клапан күйі 100-ден төмен болған цикл қабылданды, сондықтан барлық көрсеткіш осы бинарлық анықтамаға ғана қатысты. Сапа Precision, Recall және F1-score арқылы бағаланды [7]:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

мұндағы TP — дұрыс анықталған ауытқулар, FP — жалған дабылдар, FN — өткізіп алынған ауытқулар. Precision белгіленген циклдердің қаншасы шынымен ауытқу екенін, Recall нақты ауытқулардың қаншасы табылғанын көрсетеді; F1-score осы екеуінің гармониялық орташасы.

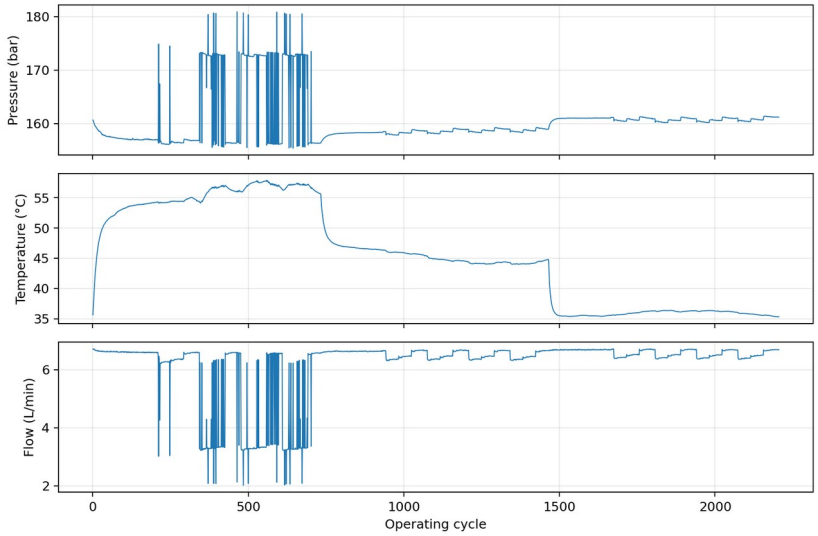
Бұл бір деректер жиынтығында кездейсоқ бөлінген үлгінің нәтижесі екенін ескеру керек. Іргелес циклдер бір-біріне өте ұқсас болуы, ал басқа компоненттердің күйі қатар өзгеруі мүмкін, сондықтан метрикалар шынайы жағдайдан жақсырақ көрінуі ықтимал. Тәуелсіз өндірістік объектідегі дәлдікпен тікелей салыстыруға болмайды [2; 3].

Isolation Forest үшін ағаштар саны 100, random_state = 42 және contamination = 0,05 қабылданды [4]. One-Class SVM қалыпты бақылаулар жиынының шекарасын ядролық түрлендіру арқылы бағалайды: біз радиалдық базистік функциялы (RBF) ядроны, nu = 0,05 және gamma = scale параметрлерін қолдандық, ал жаңа бақылау осы шекарадан тыс қалса, ауытқу деп есептелді [8]. Local Outlier Factor (LOF) нүктенің жергілікті тығыздығын көршілерінің тығыздығымен салыстырады: көршілерінен сирек орналасқан нүктелер ерекше деп саналады. Тест бақылауларын бағалау үшін novelty = True және n_neighbors = 20 параметрлері қойылды [9].

Үш алгоритм де бірдей деректерде және бірдей оқыту–валидация–тест бөлінісінде салыстырылды. StandardScaler тек оқыту деректері бойынша бейімделді, ал әр модельдің шегі валидацияда таңдалып, тест кезінде өзгертілмеді. Гиперпараметрлерді толық оңтайландыру жүргізілген жоқ.

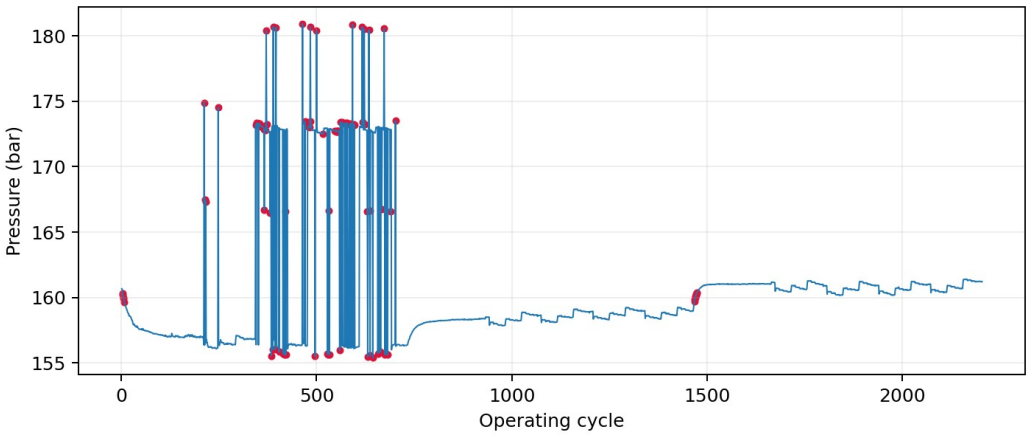
Нәтижелер және талқылау

Алдымен деректердің құрылымын қарап шықтық. Қысым мен шығын шамамен 350–700-ші циклдер аралығында күрт ауытқиды, ал температура баяу, ұзағырақ уақыт ауқымында өзгереді (1-сурет). Бұл режимдердің біркелкі еместігін көрсетеді, бірақ ақаудың себебін өздігінен айтып бере алмайды.



Сурет 1 - 2205 цикл бойынша қысым, температура және шығынның орташа мәндері

Барлық 2205 циклде оқытылған бастапқы Isolation Forest моделі 111 циклді (5,03%) аномалиялық, 2094 циклді қалыпты деп белгіледі (2-сурет). Бұл үлестің contamination = 0,05 параметрімен берілген күтілетін деңгейге жақын болуы заңды, сондықтан оны диагностикалық дәлдік ретінде қабылдауға болмайды.



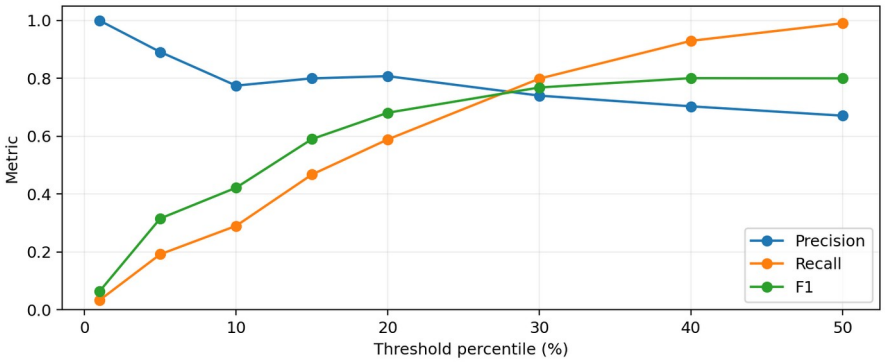
Сурет 2 - Қысым қатарындағы Isolation Forest белгілеген аномалиялар

Клапан күйінің деңгейі төмендеген сайын аномалия деп белгіленген циклдердің үлесі 3,47%-дан 8,06%-ға дейін өсті (2-кесте). Бұл бақыланған статистикалық байланыс қана: клапан күйінің әсерін басқа компоненттердегі қатар өзгерістерден бөлетін арнайы тәжірибе жасалмағандықтан, себеп-салдарлық қорытынды шығаруға болмайды.

Кесте 2 – Клапан күйі бойынша аномалиялардың үлестірілуі

Клапан күйі	Цикл саны	Аномалия	Үлесі, %
100	1125	39	3,47
90	360	18	5,00
80	360	25	6,94
73	360	29	8,06

Валидацияда F1-score ең жоғары болған шек — 40-процентиль (3-сурет). Осы шекпен тәуелсіз сақталған 441 тест циклінде Precision = 0,7207, Recall = 0,9207 және F1-score = 0,8085 алынды (3-кесте). Бұл сандар клапан күйінің ауытқуын бинарлық белгілеуге қатысты, сорғының ішкі ағып кетуіне емес.

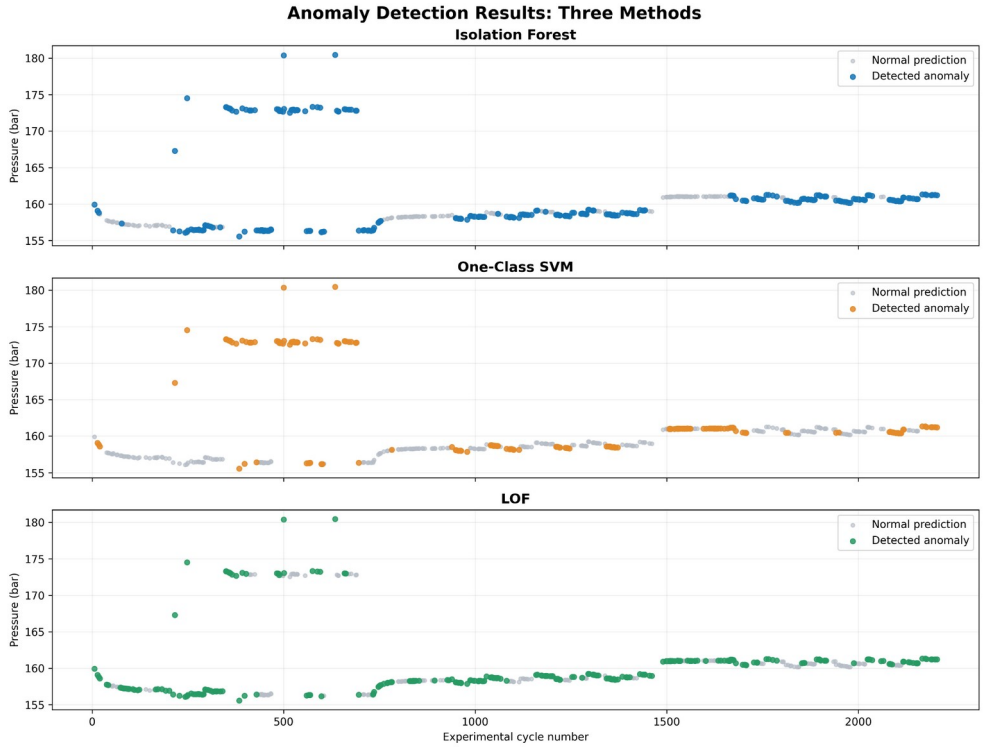


Сурет 3 - Валидация жиынтығында шешім шегінің метрикаларға әсері

Кесте 3 – Тест жиынтығы бойынша қорытынды көрсеткіштер

Көрсеткіш	**Мән**
Тест циклдері	441
TP / FP / TN / FN	209 / 81 / 133 / 18
Precision	72,07%
Recall	92,07%
F1-score	80,85%

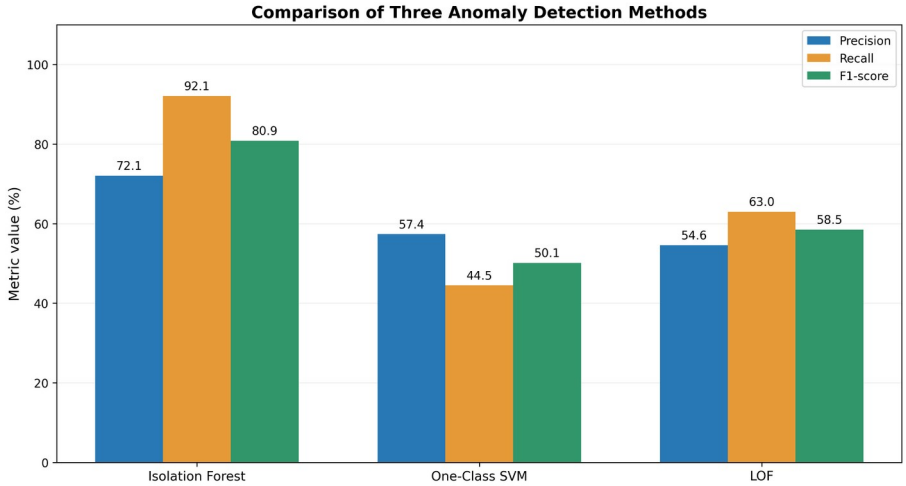
Тест жиынтығындағы бірдей 441 цикл үшін үш модельдің болжамдары қысымның циклдік орташа мәнімен қатар салыстырылды (4-сурет). Әдістердің барлығы бірдей көлденең және тік масштабта берілген; түрлі-түсті нүктелер модель аномалия деп белгілеген, сұр нүктелер қалыпты деп санаған циклдер. Қысым осі тек көрнекілік үшін алынды: әр модель қысыммен бірге температура мен шығынды да ескереді. Барлық метрикалар мен болжамдар осы жұмыста Python ортасында есептелген, олар басқа мақаладан алынған жоқ.



Сурет 4 - Үш әдіс бойынша аномалиялық циклдерді анықтау нәтижелері

4-суреттен әртүрлі әдістердің бірдей қысым аймақтарын әртүрлі жіктейтіні көрінеді. Бұл айырмашылық алгоритмдердің аномалияны бағалау қағидаларына және таңдалған шектерге байланысты. Боялған нүктелерді нақты техникалық ақаудың дәлелі деп санауға

болмайды: сапа тест жиынтығындағы мақсатты белгілермен есептелді. Метрикалардың жиынтық салыстыруы 5-суретте және 4-кестеде берілген.



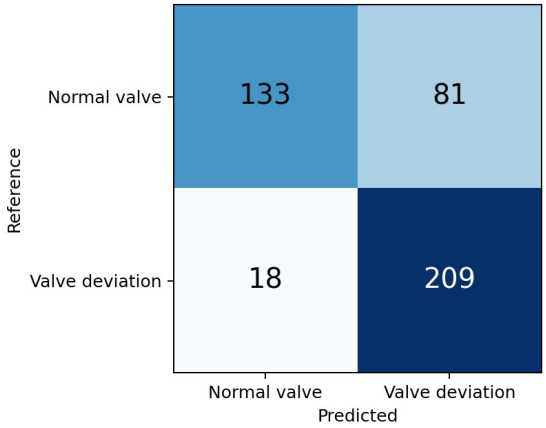
Сурет 5 - Үш әдістің Precision, Recall және F1-score көрсеткіштерін салыстыру

Кесте 4 – Үш әдістің тестілік көрсеткіштері

Әдіс	Шек, процентиль	Precision, %	Recall, %	F1-score, %
Isolation Forest	40	72,07	92,07	80,85
One-Class SVM	30	57,39	44,49	50,12
LOF	50	54,58	63,00	58,49

4-кесте бойынша Isolation Forest F1-score (80,85%) жағынан ең жақсы нәтиже берді: ол LOF-тан 22,36, One-Class SVM-нен 30,73 пайыздық тармаққа жоғары. Оның Recall көрсеткіші де 92,07%-ға жетті. Алайда Precision 72,07% болғандықтан, жалған дабылдар әлі де айтарлықтай көп. Бұл қорытынды тек осы белгілер жиынына, таңдалған параметрлерге және осы бөліністегі деректерге қатысты. Isolation Forest артықшылығының себебін бұл жұмыста бөлек зерттеген жоқпыз; ықтимал түсіндірмелердің бірі — One-Class SVM мен LOF параметрлері (nu, gamma, n_neighbors) бекітілген мәндерде қалды, ал оларды реттеу нәтижені өзгертуі мүмкін.

Шатасу матрицасы бойынша модель клапаны өзгерген 227 циклдің 209-ын тапты, 18-ін өткізіп алды (6-сурет). Сонымен қатар қалыпты клапанмен жұмыс істеген 214 циклдің 81-і қате түрде ауытқу деп белгіленді. Демек, жоғары Recall жалған ескертулердің көбеюі есебінен алынды. Мониторинг жүйесінде мұндай нәтиже қосымша тексеруді қажет етеді, оны автоматты тоқтату командасына бірден жалғауға болмайды.



Сурет 6 - Тест жиынтығындағы шатасу матрицасы

Жұмыстың негізгі шектеулері: тек үш орташа белгі пайдаланылды; клапан белгілері сорғының ішкі ағып кетуін тікелей көрсетпейді; бөліну кездейсоқ жасалды; өлшеулер бір

стендтен алынды; басқа жабдықта жалпылау тексерілмеді. Дәл осы деректер [5] жұмысында да қолданылғандықтан, нәтижелерді салыстыру пайдалы болар еді, бірақ қойылымы (жартылай бақыланатын) мен бөлу тәсілі басқа болғандықтан, сандарды тікелей салыстыруға болмайды. Келесі қадам ретінде profile.txt үшінші бағанындағы 0/1/2 белгілері бойынша сорғының ішкі ағып кетуін жеке бағалау, деректерді эксперименттік сериялар бойынша бөлу және жалған дабылдар деңгейін төмендету жоспарланады. Әдістерді салыстыру мониторинг жүйесі үшін модельді бір ғана метрика бойынша таңдауға болмайтынын көрсетеді. Ақауды өткізіп алу қаупі жоғары болса, Recall басымдыққа ие; жалған ескертулер қымбатқа түссе, Precision-ға қосымша талап қойылады. F1-score екеуін теңгеріп бағалауға көмектеседі, бірақ нақты объектіде шығын функциясы мен қауіпсіздік талаптарын бөлек белгілеу керек [2; 3].

Жүйені басқару модуліне енгізгенде алгоритмнің шығысын тікелей апаттық тоқтату командасына айналдыруға болмайды. Бастапқы кезеңде аномалия туралы хабарлама операторға және трендтерді тіркеу жүйесіне беріледі; тұрақты ауытқу расталғанда ғана регламентке сәйкес қосымша әрекеттер қарастырылады. Келесі зерттеуде сорғының ішкі ағып кету белгілерімен жеке тексеру, уақыт бойынша бөлінген сынақ, гиперпараметрлерді тең дәрежеде оңтайландыру және бірнеше іске қосу бойынша нәтижелердің орнықтылығын бағалау жоспарланады.

Қорытынды

Гидравликалық стендтің 2205 жұмыс циклі бойынша қысым, температура және шығын көрсеткіштерін пайдаланып, Isolation Forest, One-Class SVM және LOF әдістері салыстырылды. Клапаны қалыпты күйдегі деректермен оқытылып, шегі валидацияда таңдалған модельдер бірдей 441 тест циклінде тексерілді. Isolation Forest Precision 72,07%, Recall 92,07% және F1-score 80,85% көрсетті; One-Class SVM үшін F1-score 50,12%, LOF үшін 58,49% болды. Осы экспериментте Isolation Forest үш әдістің ішінде ең жоғары F1-score нәтижесіне жетті.

Жұмыс машиналық оқытуға негізделген мониторингтің мүмкіндігін және валидациялық шекті дұрыс таңдаудың маңызын көрсетті. Алайда нәтижелер клапан күйіне қатысты; сорғының ішкі ағып кетуін анықтау сапасы бөлек бағалануы керек. Болашақта сорғы белгісін пайдалану, уақыттық белгілерді қосу, тәуелсіз сынақ жүргізу, жалған дабылдарды азайту және ескерту мен басқару әрекеттерін қауіпсіз логикамен байланыстыру қажет.

Әдебиеттер тізімі:

1. ISO 17359:2018. Condition monitoring and diagnostics of machines — General guidelines. – Geneva: International Organization for Standardization, 2018.

2. Aggarwal C. C. Outlier Analysis. – 2nd ed. – Cham: Springer, 2017. – 466 p. – DOI: 10.1007/978-3-319-47578-3.

3. Pang G. et al. Deep learning for anomaly detection: A review //ACM computing surveys (CSUR). – 2021. – Т. 54. – №. 2. – С. 1-38.

4. Liu F. T., Ting K. M., Zhou Z. H. Isolation forest //2008 eighth ieee international conference on data mining. – Ieee, 2008. – С. 413-422.

5. Dong Y., Chen K., Ma Z. Comparative study on semi-supervised learning applied for anomaly detection in hydraulic condition monitoring system //2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). – IEEE, 2023. – С. 1702-1708.

6. Helwig N., Pignanelli E., Schütze A. Condition Monitoring of Hydraulic Systems [Data set]. – UCI Machine Learning Repository, 2015. – DOI: 10.24432/C5CW21.

7. Fabian P. Scikit-learn: Machine learning in Python //Journal of machine learning research. – 2011. – Т. 12. – С. 2825.

8. Schölkopf B. et al. Estimating the support of a high-dimensional distribution //Neural computation. – 2001. – Т. 13. – №. 7. – С. 1443-1471.

9. Breunig M. M. et al. LOF: identifying density-based local outliers //Proceedings of the 2000 ACM SIGMOD international conference on Management of data. – 2000. – С. 93-104.