Міністерство освіти і науки України

Харківський національний університет радіоелектроніки

Факультет комп’ютерної інженерії та управління
(повна назва)

Кафедра електронних обчислювальних машин
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

Рівень вищої освіти другий (магістерський)

Методи програмного моніторингу технічного обслуговування на елеваторному комплексі

(тема)

Виконав:

студент II курсу, групи СПм-23-2

Мостовий А.В.
(прізвище, ініціали)

Спеціальність

123 «Комп’ютерна інженерія»
(код і повна назва спеціальності)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма

Системне програмування
(повна назва освітньої програми)

Керівник: доц. Піскарьов О.М.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри ЕОМ _____ Коваленко А.А.
(підпис) (прізвище, ініціали)

2025 р.

Харківський національний університет радіоелектроніки

Факультет комп'ютерної інженерії та управління

Кафедра електронних обчислювальних машин

Рівень вищої освіти другий (магістерський)

Спеціальність 123 «Комп'ютерна інженерія»
(код і повна назва)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма Системне програмування
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
"____" _____ 20__ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенту Мостовому Артему Віталійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Методи програмного моніторингу технічного обслуговування на елеваторному комплексі

затверджена наказом по університету від " 22 " листопада 2024 р. № 1236 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 20 січня 2025 р.

3. Вхідні дані до роботи методи прогнозування, критерії якості методів аналізу даних, методи первинної обробки даних, технології та платформи системи збору та аналізу даних, засоби впровадження методів програмного моніторингу, опис експериментальних даних, оцінка ефективності, аналіз результатів тестування

4. Перелік питань, що потрібно опрацювати у роботі

1) огляд існуючих методів аналізу даних
2) розробка архітектури системи збору та аналізу даних
3) впровадження методів програмного моніторингу
4) дослідження методів програмного моніторингу
5) висновки

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) 22 слайди

6. Консультанти розділів роботи (заповнюється за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Огляд існуючих методів аналізу даних	01.10.24-19.10.24	
2	Розробка архітектури системи збору та	21.10.24-10.11.24	
	аналізу даних		
3	Впровадження методів програмного	12.11.23-21.11.24	
	моніторингу		
4	Дослідження методів програмного моніторингу	23.11.24-18.12.24	
5	Оформлення матеріалів кваліфікаційної роботи	20.12.24-02.01.25	
6	Подання кваліфікаційної роботи керівникові та	04.01.25-08.01.25	
	її попередній захист		
7	Подання кваліфікаційної роботи на	09.01.25-14.01.25	
	рецензування		

Дата видачі завдання 22 листопада 2024 р.

Студент _____
(підпис)

Керівник роботи _____ доц. Піскарьов О.М.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 95 с., 24 рис., 2 дод., 20 джерел.

МАШИННЕ НАВЧАННЯ, ПРОГНОЗУВАННЯ ПОЛОМОК, МОДЕЛІ АНАЛІЗУ ДАНИХ, ОБРОБКА ДАНИХ, АРХІТЕКТУРА СИСТЕМИ, ОПТИМІЗАЦІЯ ОБСЛУГОВУВАННЯ, НАДІЙНІСТЬ ОБЛАДНАННЯ

Метою кваліфікаційної роботи є розробка методів програмного моніторингу технічного обслуговування обладнання елеваторного комплексу, спрямованих на підвищення надійності обладнання, оптимізацію процесів обслуговування та запобігання аварійним ситуаціям.

У ході виконання кваліфікаційної роботи було проведено аналіз існуючих методів прогнозування технічного стану обладнання, визначено їх ефективність та обґрунтовано вибір моделей машинного навчання. На основі цього аналізу розроблено архітектуру системи збору, зберігання та аналізу даних, що дозволяє інтегрувати інформацію з різних джерел.

У роботі розроблено та реалізовано методи моніторингу ТО на базі машинного навчання. Проведено збір та підготовку реальних даних елеваторного комплексу, які використовувались для навчання та тестування моделей. В ході досліджень оцінено точність та стабільність розроблених методів, а також підтверджено їх ефективність.

Результати роботи свідчать про високу ефективність впроваджених методів програмного моніторингу. Запропонована система має потенціал для практичного застосування у виробничих процесах елеваторного комплексу, сприяючи підвищенню загальної ефективності роботи обладнання та зниженню витрат на його обслуговування.

ABSTRACT

Master’s thesis: 95 pages, 24 figures, 2 appendices, 20 sources.

MACHINE LEARNING, FAILURE PREDICTION, DATA ANALYSIS MODELS, DATA PROCESSING, SYSTEM ARCHITECTURE, SERVICE OPTIMIZATION, EQUIPMENT RELIABILITY

The major goal of this thesis is to develop methods for software monitoring of maintenance of elevator complex equipment aimed at improving equipment reliability, optimizing maintenance processes and preventing emergencies.

In the course of this qualification work, an analysis was conducted on existing methods for predicting the technical condition of equipment. The effectiveness of these methods was assessed, and a justification for the selection of machine learning models was provided. Based on this analysis, an architecture for a data collection, storage, and analysis system was developed, enabling the integration of information from various sources.

Machine learning-based monitoring methods were created and implemented. Real data from an elevator complex was collected and prepared for use in training and testing the models. The research evaluated the accuracy and stability of the developed methods and confirmed their effectiveness.

The results indicate that the implemented software monitoring methods are highly efficient. The proposed system has the potential for practical application in the production processes of the elevator complex, contributing to an increase in overall equipment efficiency and a reduction in maintenance costs.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ПЛК – програмований логічний контролер

СУБД – система управління базами даних

ТО – технічне обслуговування

AUC – Area Under the Curve

FN – False Negative

FP – False Positive

FPR – False Positive Rate

GC – Google Colab

GCP – Google Cloud Platform

GRU – Gated Recurrent Unit

LSTM – Long Short-Term Memory

ODBC – Open Database Connectivity

ROC – Receiver Operating Characteristic

ROCC – Receiver Operating Characteristic curve

SCADA – Supervisory Control And Data Acquisition

SVM – Support Vector Machine

TN – True Negative

TPR – True Positive Rate

ВСТУП

За останні роки дедалі більше підприємств у різних галузях промисловості активно інтегрують системи збору даних, що базуються на використанні датчиків та технологій Інтернету речей (IoT). Такі системи забезпечують отримання інформації про функціонування технологічного обладнання в режимі реального часу: температура, рівень вібрації, електричні параметри та інші. Однак, значні обсяги цих даних залишаються недостатньо проаналізованими та не застосовуються у повному обсязі для вирішення практичних задач, що обмежує їх потенціал. Одним з найважливіших напрямків використання цих даних є впровадження систем прогнозування технічного обслуговування (ТО) обладнання. Завдяки сучасним методам машинного навчання та аналізу даних, на основі таких записів можливо створювати високоточні математичні моделі, що дозволяють прогнозувати моменти можливих відмов обладнання. Це надає підприємствам змогу завчасно планувати технічне обслуговування, що значно зменшує ризики раптових поломок та незапланованих зупинок виробничих процесів. Впровадження таких методів має стратегічне значення для підприємств, оскільки воно дозволяє змінити підхід до ТО – від реактивного до проактивного. Завдяки цьому суттєво скорочуються простої обладнання, знижуються витрати на позапланові ремонти, підвищується загальна ефективність виробництва, а також зростає надійність технологічних
процесів [1, 2].

Обрана тема дослідження має достатню актуальність, оскільки вона гармонійно вписується у світові тенденції розвитку промисловості та відповідає принципам концепції «Індустрія 4.0». Розробка інноваційних рішень у цій сфері має широке практичне застосування й здатна забезпечити значні економічні переваги для промислових підприємств [3].

1 ОГЛЯД ІСНУЮЧИХ МЕТОДІВ АНАЛІЗУ ДАНИХ

У наш час, коли дані стають невід'ємною частиною всієї діяльності, вміння ефективно аналізувати та отримувати справжню цінність від цифрової інформації є важливим фактором для досягнення успіху для бізнесу [3]. Попри величезний обсяг даних, що генеруються щодня, лише 0,5% з них піддаються аналізу й використовуються для виявлення, покращення та дослідження. Хоча цей відсоток здається незначним у контексті обсягу доступної цифрової інформації, навіть така мала частка становить велику кількість даних.

У такій ситуації, коли обсяг даних є величезним, а час обмежений, знання про те, як збирати, організовувати, управляти та осмислювати цю, потенційно цінну для бізнесу інформацію, стає важливим викликом. У той же час, аналіз даних є ефективним інструментом для вирішення цього завдання. У науковому контексті аналіз даних застосовує складні підходи та сучасні методи для дослідження та експериментів з даними. Натомість у бізнесі ці дані використовуються для прийняття рішень, що дозволяють покращити ефективність організації [4].

Аналіз даних являє собою процес збору, моделювання та обробки інформації за допомогою різноманітних статистичних та логічних методів. Компанії активно використовують аналітичні інструменти та процеси для отримання знань, які є основою для прийняття як стратегічних, так й операційних рішень. Усі ці методи значною мірою ґрунтуються на двох основних підходах: кількісних та якісних. Окрім цих категорій, існують й інші типи даних, які необхідно враховувати перед тим, як здійснити поділ на складні процеси аналізу даних [5].

До таких категорій належать Big data (великі дані), що представляють собою великі набори інформації, які потрібно аналізувати за допомогою сучасних програмних засобів для виявлення закономірностей та тенденцій.

Вони є одними з найбільш цінних аналітичних активів, оскільки дозволяють обробляти великі обсяги даних значно швидше.

Метадані, в свою чергу, є даними, що описують інші дані, узагальнюючи ключову інформацію про них й полегшуючи їх пошук та повторне використання для подальших цілей. Дані в реальному часі одразу доступні після отримання, що робить їх надзвичайно корисними для прийняття рішень на основі актуальної інформації. Окрім того, існують й машино-згенеровані дані, які утворюються різними пристроями, такими як комп'ютери, телефони чи веб-сайти, без участі людини.

Коли йдеться про аналіз даних, існує визначений порядок дій [5], якого слід дотримуватися для досягнення необхідних результатів. Процес аналізу можна поділити на кілька ключових етапів. Спочатку потрібно чітко визначити цілі аналізу, наприклад, спрогнозувати, коли обладнання потребуватиме ремонту або заміни, визначити, які компоненти найбільш схильні до поломки, або виявити фактори, що впливають на термін служби обладнання. Коли цілі окреслені, наступним кроком є збір необхідних даних. Це можуть бути історії ремонтів, дані про використання обладнання, інформація про виробничі процеси, тощо. Наступним етапом є очищення даних - необхідно обробити отриману інформацію, усунувши дублікатні записи, помилки у форматуванні та непотрібні елементи. Після цього дані будуть готові для подальшого аналізу. Важливим етапом є безпосередньо сам аналіз, де використовуються різноманітні методи, такі як статистичні методи, регресійний аналіз, нейронні мережі або аналіз для виявлення тенденцій, кореляцій та закономірностей. Це дозволяє отримати відповіді на питання, визначені на початковому етапі. Завершальний етап полягає в інтерпретації результатів аналізу, де на основі отриманих даних розробляються плани дій. Це може включати планування ремонту обладнання, розробку заходів щодо обслуговування або оптимізацію виробничих процесів. Також на цьому етапі може бути корисно виявити обмеження аналізу та попрацювати над їх усуненням.

1.1 Аналіз існуючих методів прогнозування

Логістична регресія є контрольованим алгоритмом машинного навчання, що використовується для вирішення задач класифікації. Основна мета цього підходу полягає у прогнозуванні ймовірності належності об'єкта до певного класу [6]. Він базується на статистичному аналізі взаємозв’язку між незалежними змінними та залежною змінною, що зазвичай має бінарний характер. Завдяки своїй здатності моделювати ймовірності, логістична регресія є ефективним інструментом для ухвалення обґрунтованих рішень у різних сферах, включаючи медицину, фінанси, маркетинг та інші галузі. Логістична регресія призначена для прогнозування категоріальної залежної змінної, тому її результати мають дискретний характер. Наприклад, значення можуть бути представлені як «так» або «ні», «0» або «1», «правда» або «брехня». Однак замість видачі точних значень «0» і «1», модель обчислює ймовірності, які лежать у межах від 0 до 1. Цей алгоритм багато в чому подібний до лінійної регресії, але відрізняється своїм призначенням. Лінійна регресія використовується для вирішення задач прогнозування (регресії), тоді як логістична регресія застосовується для класифікації. У логістичній регресії замість апроксимації лінії регресії використовується логістична функція, яка утворює S-подібну криву. Ця крива дозволяє моделювати ймовірності двох класів, наприклад, визначати, чи є клітини раковими, чи ні, або оцінювати, чи страждає миша ожирінням, виходячи з її ваги.

Логістична регресія є потужним інструментом машинного навчання, оскільки вона здатна класифікувати нові дані, працюючи з безперервними та дискретними наборами даних. Цей алгоритм дозволяє не лише передбачати класи, але й оцінювати ймовірність належності об'єкта до певного класу. Вона також може допомогти виявити найважливіші змінні, які впливають на класифікацію, що робить її особливо цінною в різних аналітичних задачах.

Логістична регресійна модель використовує логістичну функцію для перетворення вихідних значень лінійної регресійної функції з неперервного

діапазону на категоріальні значення. Ця функція, також відома як сигмоїдальна функція, що демонструє дійсний набір вхідних незалежних змінних у значення від 0 до 1. Завдяки цьому підходу модель здатна прогнозувати ймовірності та виконувати класифікацію, розподіляючи об'єкти до одного з двох можливих класів [6].

Незалежні вхідні ознаки *X* дорівнюють:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}, \quad (1.1)$$

а залежною змінною є Y, яка приймає лише двійкове значення («0» або «1»):

$$Y = \begin{cases} 0 \ \ if \ Class \ 1 \\ 1 \ \ if \ Class \ 2 \end{cases}. \quad (1.2)$$

Далі необхідно використовувати «мультилінійну» функцію z до вхідних змінних x:

$$z = (\sum_{i=1}^{n} w_i x_i) + b. \quad (1.3)$$

У логістичній регресії кожне спостереження x_i з множини вхідних ознак

X асоціюється з набором вагових коефіцієнтів w, а b - виступає як зміщення або перехоплення. Це можна представити у вигляді скалярного добутку ваги та зміщення.

$$z = w_1 x_1 + w_2 x_2 + \cdots \ w_n x_n + b = w \cdot X + b. \quad (1.4)$$

На цьому етапі модель застосовує сигмоїдальну функцію, де вхідним значенням є z , а вихідне значення представляє ймовірність, що лежить у

діапазоні від 0 до 1 [6]. Таким чином, передбачуване обчислюється за допомогою формули:

$$\sigma(z) = \frac{1}{1+e^{-t}}. \quad (1.5)$$

Сигмовидна функція перетворює дані неперервної змінної в ймовірність, тобто від 0 до 1 (рисунок 1.1).

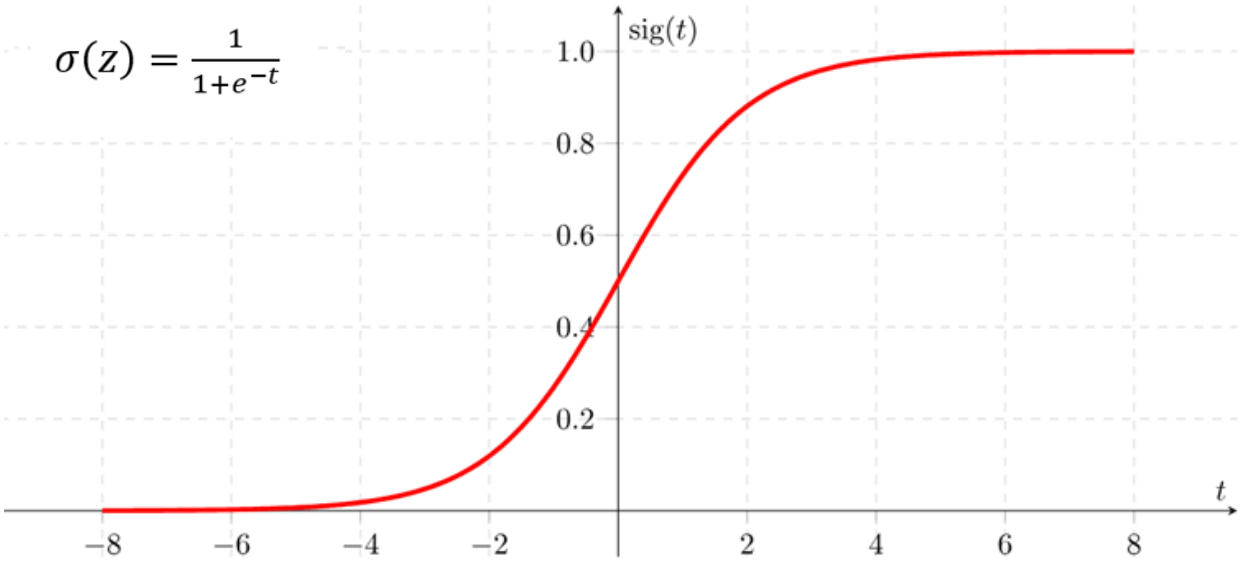


Рисунок 1.1 – Вигляд сигмовидної функції

Логістична регресія базується на кількох ключових припущеннях. По-перше, кожне спостереження має бути незалежним від інших, це означає, що немає кореляції між будь-якими вхідними змінними. По-друге, залежна змінна повинна бути двійковою, тобто приймати лише два можливі значення. Якщо задача передбачає більше ніж два класи, то для таких випадків використовуються модифікації, наприклад функція softmax. Крім того, між незалежними змінними та логарифмічними коефіцієнтами залежної змінної повинна існувати лінійна залежність. Інше важливе припущення полягає в тому, що у наборі даних не повинно бути викидів, оскільки вони можуть

вплинути на точність моделі. Для ефективної роботи моделі важливо, щоб розмір вибірки був достатньо великим, оскільки це дозволяє отримати надійні та стабільні результати [6].

Дерево рішень є одним з найпопулярніших та ефективних інструментів у алгоритмах контрольованого навчання, що широко використовуються як для задач класифікації, так і для задач регресії [7]. Основна ідея методу полягає у побудові дерева (рисунок 1.2), яке служить моделлю для прийняття рішень на основі атрибутів даних. Дерево рішень представляє собою деревоподібну структуру, де кожен вузол дерева (внутрішній вузол) відповідає тесту або порівнянню одного з атрибутів вхідних даних. Гілки, які виходять із цього вузла, відображають різні можливі результати тесту або значення атрибуту. Термінальні (листові) вузли містять кінцеві рішення або класифікаційні мітки, тобто відповідь або прогноз для певного спостереження на основі його характеристик.

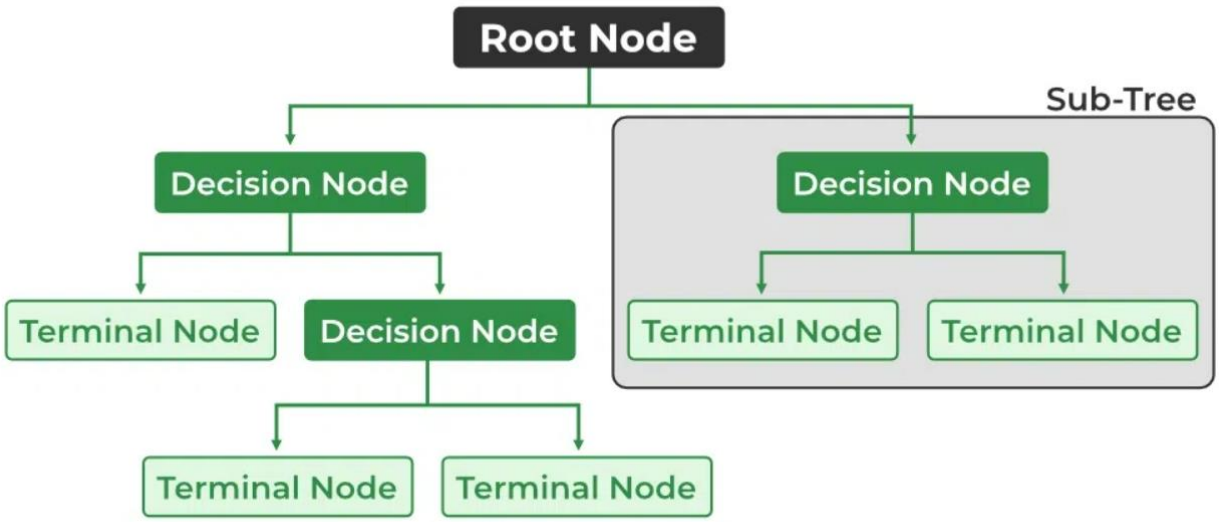


Рисунок 1.2 - Дерево рішень

Процес побудови дерева рішень починається з кореневого вузла, де здійснюється тестування на найбільш значущому атрибуті. Далі цей процес

повторюється рекурсивно для кожного підвузла: дані поділяються на підмножини на основі результатів попереднього тесту, що дозволяє поступово уточнювати прогноз для кожного конкретного випадку. Розбиття триває до досягнення певних критеріїв зупинки, які можуть включати досягнення максимальної глибини дерева, мінімальної кількості вибірок для подальшого поділу вузла, або до того, як всі елементи в підмножині будуть належати одному класу, що необхідно для класифікації.

Однією із головних переваг дерева рішень є його інтерпретованість та простота у використанні. Кожен крок прийняття рішення можна відслідкувати, що робить модель зрозумілою та доступною для подальшого аналізу. Це дозволяє не лише будувати ефективні прогнози, але й надавати чітке пояснення того, чому було прийнято те чи інше рішення.

Загалом дерево рішень є дуже гнучким методом, здатним працювати з різноманітними типами даних і вирішувати завдання, де важливою є не тільки точність прогнозу, але й можливість пояснити, як до нього дійшли. Тим не менш, дерево рішень може схильне до перенавчання (overfitting), особливо при занадто великій глибині дерева, тому часто використовуються різні методи оптимізації, такі як обмеження глибини дерева або використання ансамблів дерев, таких як випадковий ліс (random forest), для покращення загальної продуктивності моделі [7].

Під час навчання алгоритм дерева рішень обирає найкращий атрибут для поділу даних на основі показників, таких як ентропія або домішка Джині, які оцінюють рівень невизначеності чи випадковості в підмножинах даних. Метою є вибір атрибута, який дозволяє максимізувати приріст інформації або мінімізувати домішку після розподілу, що забезпечує більш чітке розмежування між класами. Вибір атрибутів зазвичай оцінюється за допомогою спеціальної міри, відомої як міра вибору атрибутів, що допомагає визначити, який атрибут створить найбільш однорідні підмножини, підвищуючи таким чином точність моделі.

Цей процес розбиття повторюється рекурсивно для кожної

підмножини, доки не буде досягнуто зупинки — або коли всі елементи в підмножині належать одному класу, або коли подальше розбиття не дає суттєвого покращення точності. Процес розбиття називається рекурсивним розбиттям, і він є основним етапом у побудові дерева рішень.

Побудова класифікатора дерева рішень не вимагає спеціальних знань про предметну область або налаштувань параметрів, що робить цей метод ідеальним для дослідницьких завдань та виявлення нових знань. Дерева рішень також можуть ефективно обробляти дані з високою розмірністю, що дозволяє використовувати цей підхід для роботи з великими та складними наборами даних.

Алгоритм дерева класифікації та регресії працює шляхом послідовного розбиття даних у кожному вузлі дерева на основі значень атрибутів [8], щоб досягти оптимального поділу для класифікації або регресії. Якщо дані в вузлі m складаються з множини зразків Q_m та мають n_m зразків, то алгоритм класифікації та регресії виглядає наступним чином:

$$G(Q_m, t_m) = \frac{n_m^{left}}{n_m} H\left(Q_m^{left}(t_m)\right) + \frac{n_m^{right}}{n_m} H\left(Q_m^{right}(t_m)\right), \quad (1.6)$$

де H - міра домішок лівої та правої підмножин у вузлі m (це може бути ентропія або домішка Джині);

n_m — кількість екземплярів у лівій та правій підмножинах у вузлі m.

Для вибору параметру можемо таким чином:

$$t_m = t_m\left(Q_m, t_m\right). \quad (1.7)$$

Випадковий ліс, або Random Forest – це потужний ансамблевий метод машинного навчання, який застосовується для класифікації, регресії та інших задач [8]. Основною ідеєю алгоритму є побудова великої кількості дерев рішень під час тренування моделі, що дозволяє отримати більш точні

прогнози. В результаті, алгоритм формує остаточне рішення, яке є або модою (для класифікації), або середнім прогнозом (для регресії) з усіх побудованих дерев. Використання багатьох дерев дозволяє зменшити варіативність та підвищити точність моделі, однак іноді алгоритм схильний до перенавчання, якщо не застосовуються відповідні стратегії для уникнення цього.

Розширення та вдосконалення методу було запропоновано Лео Брейманом і Аделем Катлером, а сам алгоритм «Random Forests» став їх торговою маркою. Він поєднує дві ключові ідеї: метод Беггінга «Bootstrap Aggregating», запропонований Брейманом, дозволяє створювати різні варіанти дерев, використовуючи випадкові підмножини даних для тренування, а метод випадкових підпросторів «Random Subspaces», запропонований Tin Kam Ho, зменшує кореляцію між деревами шляхом випадкового вибору підмножин ознак для кожного дерева. Це поєднання дозволяє досягати високої стабільності та точності в результатах моделі, що робить Random Forest одним з найбільш ефективних алгоритмів для багатьох типів задач машинного навчання [8].

Алгоритм навчання класифікаційного моделера для випадкового лісу працює таким чином: на початковому етапі з навчальної вибірки генерується випадкова підвибірка з розміром n, при цьому деякі приклади можуть потрапити в підвибірку декілька разів, а інші — не потрапити зовсім. Це дозволяє створити варіативність, оскільки приблизно $N/3$ прикладів залишаються не використаними для конкретного дерева. Далі для побудови кожного дерева вибирається підмножина ознак, причому з усіх доступних M ознак випадковим чином обираються m. Така випадковість допомагає уникнути сильної кореляції між деревами, що підвищує різноманітність моделі [8].

Після цього для побудови дерева рішень в кожному вузлі обирається одна з m ознак для поділу на підмножини. Для цього використовуються критерії (коефіцієнт Джині або приріст інформації), щоб вибрати ознаку, яка дасть найкраще розбиття на класи. Процес побудови дерева продовжується

до того, як всі зразки в кожному вузлі будуть належати до одного класу або поки не буде досягнуто певного критерія зупинки. Після побудови одного дерева процес повторюється: генерується нова підвибірка, випадково вибираються ознаки, і будуються наступні дерева [8]. Чим більше дерев побудовано, тим більш стабільним і точним буде класифікатор, зменшуючи ймовірність помилки на тестовій вибірці.

Класифікація об'єктів у методі випадкового лісу здійснюється шляхом голосування серед дерев, що складають модель. Кожне дерево, в залежності від набору ознак, класифікує об'єкт до одного з можливих класів. Остаточне рішення приймається на основі того, який клас отримав найбільшу кількість голосів від дерев. Це дозволяє збільшити стабільність результату і зменшити ймовірність помилки в порівнянні з одиничним деревом.

Для вибору оптимальної кількості дерев в ансамблі враховується помилка на тестовій вибірці, при цьому метою є мінімізація цієї помилки. Якщо тестова вибірка відсутня, оптимізація проводиться на основі оцінки помилки out-of-bag, яка надає змогу виміряти частку прикладів з навчальної вибірки, що були неправильно класифіковані моделлю, коли для їх класифікації не бралися до уваги дерева, що були навчені на цих прикладах. Наведений алгоритм дозволяє більш точно оцінити ефективність моделі без необхідності створення окремої тестової вибірки.

Support Vector Machine (SVM) є алгоритмом контрольованого навчання, що застосовується як для класифікації, так й для регресії, але зазвичай він більш ефективніше працює саме в задачах класифікації [9]. Головна мета SVM — знайти оптимальну гіперплощину в просторі ознак, що дозволяє ефективно розділити точки даних різних класів. Ця гіперплощина має за мету максимально збільшити відстань до найближчих точок обох класів, що дозволяє забезпечити чітке розділення. Розмірність цієї гіперплощини залежить від кількості вхідних ознак. У двовимірному просторі гіперплощина є прямою лінією, а в тривимірному просторі — площиною. Для більшої кількості ознак, коли простір стає багатовимірним,

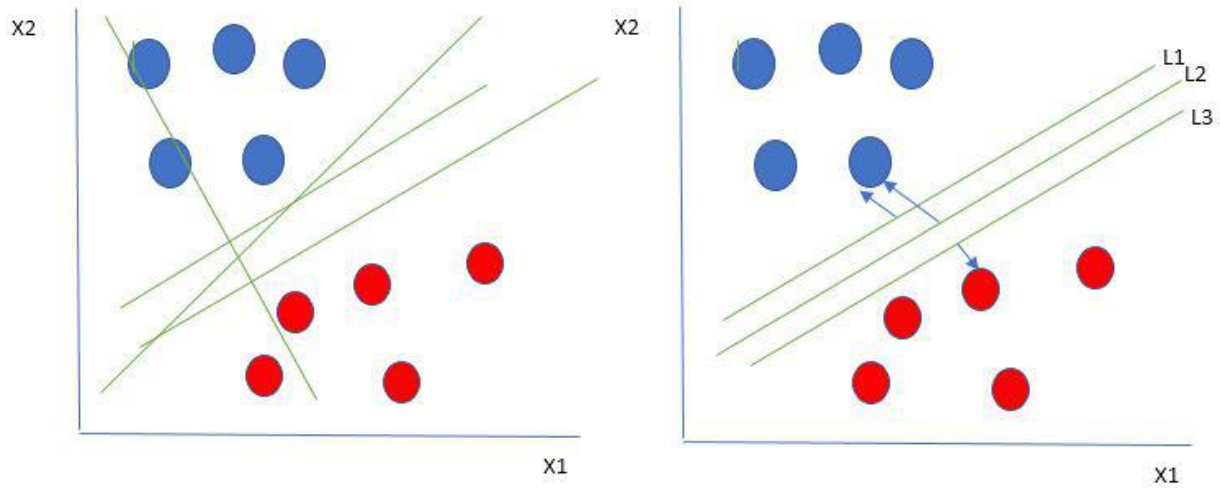
уявити гіперплощину стає складно, але алгоритм все одно працює за принципом знаходження оптимальної межі розподілу, навіть у випадку високорозмірних просторів. SVM намагається забезпечити таку структуру межі, щоб мінімізувати помилки класифікації для нових даних. У випадку, коли ми маємо дві незалежні змінні x_1 та x_2, а також одну залежну змінну, яка може приймати два можливих значення — синє коло або червоне коло, завдання класифікації полягає в тому, щоб на основі значень x_1 та x_2 віднести точку до одного з двох класів. Кожна точка з координатами (x_1, x_2) представляє певний об'єкт, а її клас (синє або червоне коло) залежить від значень цих незалежних змінних.

Алгоритм SVM, в цьому випадку, намагається знайти таку гіперплощину (в даному випадку — лінію в двовимірному просторі), яка розділяє дві категорії (сині та червоні кола) таким чином, щоб максимізувати відстань між найближчими точками кожного класу. Це дозволяє мінімізувати помилку класифікації для нових точок, яких не було в навчальній вибірці.

Таким чином, задача SVM зводиться до побудови лінії, яка максимально ефективно розділяє ці два класи в просторі. Якщо дані є лінійно роздільними, лінія може бути знайдена досить просто, однак, якщо дані не можна поділити лінійно, необхідно використовувати ядрові методи для трансформації даних в простір вищої розмірності, де лінійне розділення може стати можливим [9].

З рисунку 1.3,а зрозуміло, що є кілька ліній (в даному випадку гіперплощина тут є лінією, тому розглядаємо тільки дві вхідні характеристики x_1, x_2), що відокремлюють наші точки даних або класифікують червоні та сині кола. Одним з розумних виборів в якості найкращої гіперплощини є та, що представляє найбільшу відстань або межу між двома класами – як показано на рисунку 1.4. Якщо обираємо гіперплощину, що максимально віддалена від найближчих точок з кожного боку, то таку гіперплощину називають гіперплощиною з максимальною межею або жорсткою межею. Це означає, що відстань від цієї гіперплощини до найближчих точок класів (які

називаються опорними векторами) є найбільшою, й це забезпечує найбільшу можливу відстань між двома класами.



а) б)

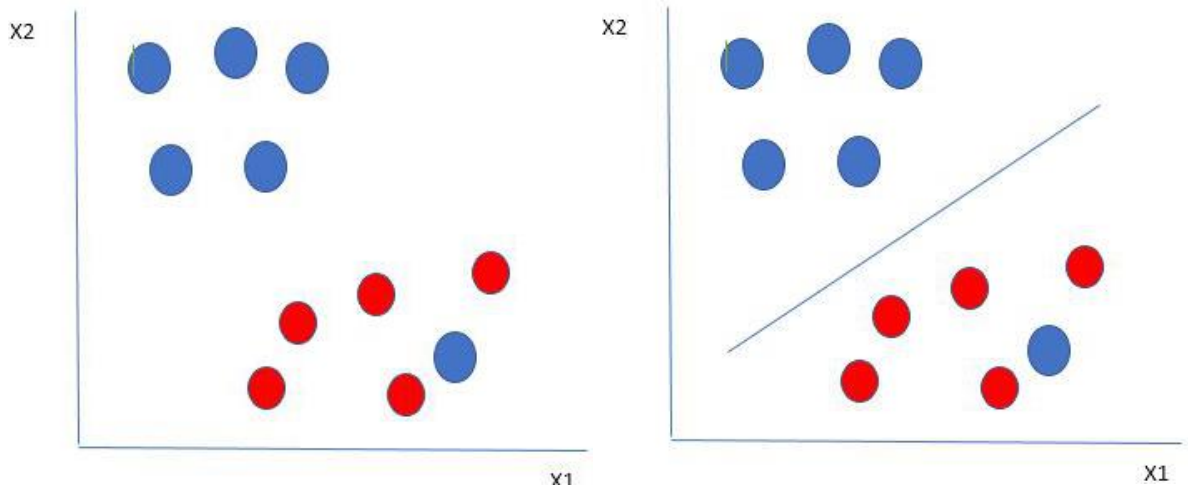
Рисунок 1.3 – Алгоритм SVM: лінійно-відокремлювані точки даних (а) та декілька гіперплощин, що відокремлюють дані від двох класів (б)

Якщо уявити на графіку (рисунок 1.3), де два класи (наприклад, сині та червоні кола) розташовані в двовимірному просторі, гіперплощина буде лінією, що розділяє ці два класи таким чином, щоб максимально відстань до найближчих точок з кожного боку була однакова. Вибір такої лінії (гіперлінії) — це основна мета SVM для лінійно роздільних даних.

На рисунку 1.3,б - якщо ми маємо кілька можливих гіперплощин для розділення даних, то вибір *L2* означає, що ми обираємо саме ту лінію, що задовольняє умову максимальної відстані до найближчих точок з обох класів. Це забезпечує оптимальне розділення і, як результат, мінімізацію помилки класифікації для нових точок. Це дозволяє алгоритму SVM мати хорошу генералізацію, оскільки така максимальна межа надає найбільшу впевненість у класифікації нових, невідомих точок, що потрапляють у область, визначену цією гіперплощиною.

Розглянемо сценарій, який показано на рисунку 1.4. У випадку, коли в

даних є викиди, як показано на рисунку 1.4,б - алгоритм SVM стає більш гнучким завдяки здатності обробляти такі ситуації. Традиційна модель SVM намагається знайти гіперплощину, яка максимізує маржу між двома класами, але коли дані містять викиди, вона може неправильно класифікувати деякі точки [9].



а) б)

Рисунок 1.4 - Алгоритм SVM: вибір гіперплощини для даних з викидом (а) та найбільш оптимізована гіперплощина (б)

Для вирішення цієї проблеми SVM використовує концепцію "м'якої межі" (soft margin). У м'якому полі, якщо точка потрапляє на неправильну сторону межі або перетинає її, додається штраф. Мета алгоритму — не тільки максимізувати маржу, але й мінімізувати штрафи за помилкові класифікації. Таким чином, замість того, щоб строго дотримуватися жорсткої межі, як це було в класичній версії SVM, алгоритм дозволяє певні порушення, але контролює їх кількість та вплив на модель.

Формула для функції втрат у випадку м'якої межі виглядає як мінімізація суми двох компонентів: першою є інверсія маржі (1/маржа), а другою — штрафи, які накладаються за порушення межі (∑штраф). Як правило, для визначення штрафів використовують втрата петлі (hinge loss),

яка має таку особливість: якщо точка не порушує межу, то штраф дорівнює нулю, але якщо точка потрапляє в неправильну область, то штраф пропорційний відстані, на яку точка перетинає межу.

Такий підхід робить SVM стійким до викидів, тому що він дозволяє класифікувати більшість точок коректно, зберігаючи високу точність, навіть якщо деякі з точок є викидами або знаходяться близько до межі.

Машини опорних векторів (SVM) можна розділити на дві основні категорії залежно від характеру межі прийняття рішень: лінійний SVM та нелінійний SVM. Лінійний SVM використовується, коли дані можна точно розділити за допомогою лінії (у 2D) або гіперплощини (у вищих вимірах). У такому випадку, межа рішень є лінійною та максимально віддаленою від найближчих точок кожного класу, що дозволяє досягти найкращої класифікації. Це найпростіший тип SVM, який добре працює, коли дані лінійно відокремлюються, тобто одна пряма або гіперплощина може повністю розділити два класи. Нелінійний SVM, у свою чергу, призначений для більш складних задач, коли дані не можна розділити лінійно. Для цього використовуються функції ядра, що перетворюють вихідні дані в простір ознак вищої розмірності, де стає можливим лінійне розділення класів. Це дозволяє побудувати складнішу нелінійну межу рішень у новому просторі ознак. Таким чином, навіть якщо дані не можна відокремити лінійно у початковому просторі, за допомогою ядра вони можуть бути розділені у більш високому вимірі, що дозволяє застосувати лінійний SVM для створення межі рішень [9].

Отже, лінійний SVM є простим та ефективним методом для лінійно роздільних даних, а нелінійний SVM, завдяки використанню функцій ядра, дозволяє працювати з більш складними, нелінійно відокремлюваними даними.

1.2 Критерії якості методів аналізу даних

Для оцінки ефективності методів аналізу даних на основі моделі машинного навчання використовується матриця помилок, яка дає змогу детально проаналізувати, як модель класифікує тестові дані. Вона містить чотири основні категорії: правильно класифіковані позитивні значення (True Positive, TP), правильно класифіковані негативні значення (True Negative, TN), неправильно класифіковані позитивні значення (False Positive, FP) і неправильно класифіковані негативні значення (False Negative, FN). Ці значення дозволяють обчислити точність моделі, яка визначається як частка правильно класифікованих значень від загальної кількості прикладів у тестовій вибірці.

У випадку бінарної класифікації матриця помилок має вигляд таблиці *2x2*, де кожен елемент таблиці вказує на кількість певних типів класифікаційних результатів. True Positive (TP) — це кількість правильно класифікованих екземплярів, що належать до позитивного класу. True Negative (TN) — це кількість правильно класифікованих екземплярів, які належать до негативного класу. False Positive (FP) — це кількість неправильно класифікованих екземплярів, які насправді належать до негативного класу, але були віднесені до позитивного. False Negative (FN) — це кількість неправильно класифікованих екземплярів, які належать до позитивного класу, але були віднесені до негативного.

Це дозволяє визначати різноманітні метрики, зокрема чутливість, специфічність, точність та інші, які дають змогу глибше зрозуміти, як модель працює. У разі багатокласової класифікації матриця помилок набуває форми *n x n*, де *n* — це кількість класів. Кожен елемент цієї матриці відображає кількість екземплярів, що були правильно чи неправильно класифіковані в контексті відповідного класу. Такий формат дозволяє отримати детальне уявлення про взаємодію моделі з кожним класом у багатокласовому середовищі.

Крива Receiver Operating Characteristic (ROC) та метрика Area Under the Curve (AUC) використовуються для оцінки ефективності класифікаційних моделей [10], зокрема для визначення їх здатності правильно відрізняти різні класи. ROC є графічним відображенням продуктивності бінарної класифікаційної моделі, де по осі X відкладається частота помилкових спрацьовувань (FPR), а по осі Y — частота істинних позитивних результатів (TPR) при різних порогах класифікації. AUC, або площа під кривою ROC, вимірює загальну ефективність моделі; значення AUC знаходиться в діапазоні від 0 до 1, де більше значення вказує на кращу здатність моделі відрізняти класи. Кінцева мета — максимізувати площу під кривою, досягнувши найвищого TPR і найнижчого FPR. На графіку, зображеному на рисунку 1.5, показано, як ці два показники взаємодіють для оцінки класифікаційної моделі.

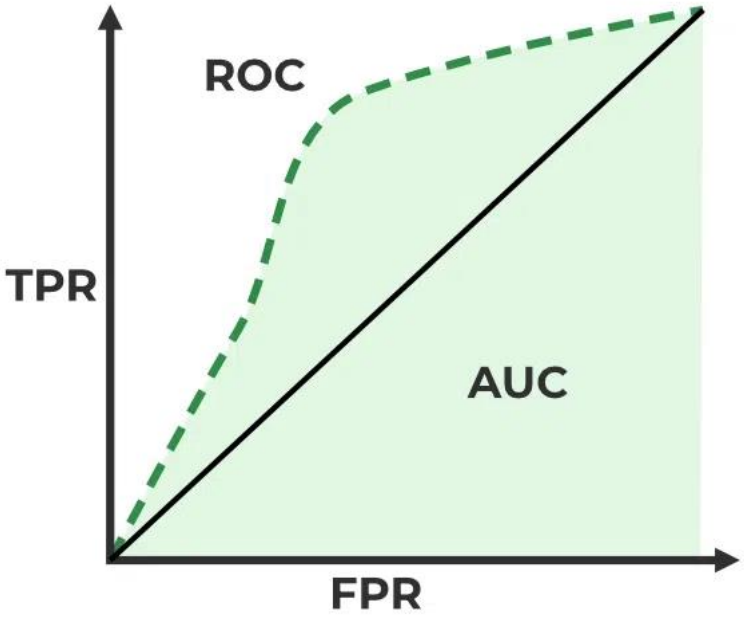


Рисунок 1.5 - Метрика оцінки класифікації ROC-AUC

Терміни True Positive Rate (TPR) та False Positive Rate (FPR) є важливими показниками для аналізу продуктивності класифікаційних

моделей через криву ROC. TPR (чутливість) визначає частку правильно класифікованих позитивних екземплярів серед усіх позитивних прикладів, що дає уявлення про здатність моделі правильно виявляти позитивні класи. З іншого боку, FPR вимірює частоту помилкових спрацьовувань, тобто частку негативних екземплярів, які були неправильно класифіковані як позитивні. Ці два показники визначають, наскільки добре модель може уникати помилок та коректно класифікувати екземпляри у різні класи. Вони базуються на матриці плутанини, де важливими компонентами є True Positive, True Negative, False Positive та False Negative, що використовуються для розрахунку різних метрик та визначення ефективності моделі, що представлено на рисунку 1.6.

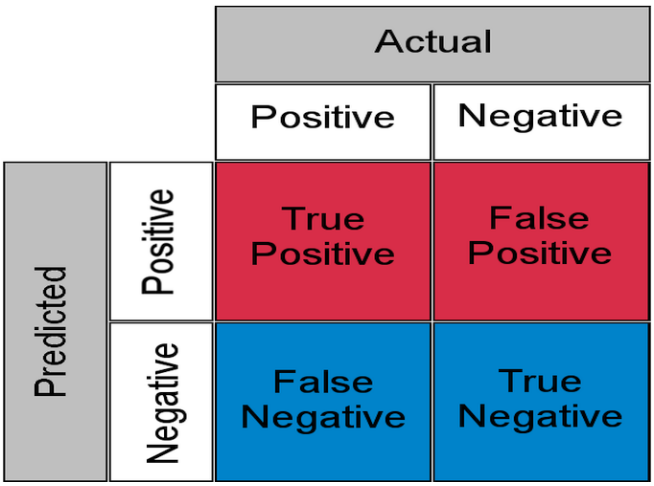


Рисунок 1.6 - Матриця плутанини для задачі класифікації

Таким чином, крива ROC та метрика AUC є важливими інструментами для оцінки продуктивності класифікаційних моделей, особливо у випадках бінарної класифікації. Крива ROC показує, як змінюється співвідношення між правильно класифікованими позитивними екземплярами (TPR) та помилковими спрацьовуваннями (FPR) при варіаціях порогу класифікації. Це дає можливість наочно оцінити здатність моделі відрізняти між двома класами при різних умовах класифікації.

Метрика AUC, що вимірює площу під кривою ROC, дає загальну оцінку ефективності моделі, де вищі значення AUC свідчать про кращу здатність моделі до класифікації. Ідеальний результат для моделі — це AUC, близьке
до 1, що вказує на те, що модель здатна максимально точно відрізняти між класами, тоді як AUC близьке до 0,5 вказує на випадкову або неефективну модель [11].

Отже, крива ROC і AUC є корисними для вибору та порівняння моделей, особливо коли є необхідність працювати з різними порогами класифікації. Вони допомагають не лише оцінити точність моделі, а й зрозуміти, як вона реагує на різні рівні ймовірності, що дозволяє регулювати її параметри для досягнення оптимальних результатів.

Коефіцієнт Джині є однією з основних метрик для оцінки якості розділення класів у задачах класифікації, зокрема в контексті дерев рішень. Цей коефіцієнт дозволяє виміряти рівень невпорядкованості або змішаності класів в конкретному вузлі моделі. Його значення варіюються від 0 до 1, де 0 свідчить про повну впорядкованість (тобто, всі екземпляри в вузлі належать до одного класу), а 1 вказує на максимальну невпорядкованість (коли екземпляри класів розподілені випадковим чином).

При оцінці ефективності моделі з низьким коефіцієнтом Джині можна сказати, що модель успішно розділяє класи, й це вказує на її ефективність. Оскільки коефіцієнт Джині орієнтований на якість розділення, він дозволяє вибирати найкращі ознаки та значення порогів для розбиття даних, максимізуючи відмінність між класами. Це робить його цінним інструментом у побудові дерев рішень, де мета полягає в тому, щоб кожен вузол максимально розрізняв класи.

Коефіцієнт Джині схожий на ентропію, яка також вимірює невпорядкованість, але між ними є різниця у тому, як вони використовуються. Обидві метрики можуть бути ефективними для класифікаційних завдань, однак вибір між ними часто залежить від

специфіки задачі. Коефіцієнт Джині має простішу форму обчислення, що робить його дещо швидшим для застосування в порівнянні з ентропією.

Таким чином, коефіцієнт Джині є важливим елементом для оцінки якості моделей класифікації, особливо в контексті дерев рішень, і допомагає виявити найбільш інформативні ознаки для класифікації.

Метрики Accuracy, Precision і Recall використовуються для оцінки продуктивності класифікаційних моделей й дають уявлення про те, наскільки добре модель виконує свою задачу класифікації.

Accuracy визначає загальний відсоток правильних прогнозів серед усіх прогнозів, включаючи як позитивні, так і негативні класи. Вона вимірює, наскільки точно модель класифікує приклади, але може бути недостатньо інформативною, якщо класи мають нерівномірне розподілення:

$$Accuracy = \frac{\text{Кількість правильних класифікацій}}{\text{Загальна кількість екземплярів}}.$$

Accuracy виражає загальний рівень вірності моделі, але вона може бути обманливою у випадках незбалансованих класів, коли один клас переважає над іншим.

Precision вимірює частку правильних позитивних прогнозів серед усіх прикладів, які модель класифікує як позитивні. Це важлива метрика, коли нас цікавить, щоб модель не давала хибнопозитивних результатів, тобто неправильно не відносила негативні екземпляри до позитивних:

$$Precision = \frac{True\ Positive}{True\ Positive\ +\ False\ Positive}.$$

Precision вимірює точність моделі у визначенні позитивних екземплярів, і вона особливо корисна, коли важливо уникнути неправильних позитивних класифікацій.

Recall або чутливість показує, скільки насправді позитивних прикладів модель змогла правильно класифікувати серед усіх позитивних прикладів у наборі даних. Високий recall означає, що модель не пропускає позитивні приклади, але може включати в прогноз і більше хибнопозитивних результатів:

$$Recall \text{ (Повнота)} = \frac{True\ Positive}{True\ Positive\ +\ False\ Negative}.$$

Recall визначає, наскільки добре модель виявляє всі фактичні позитивні екземпляри. Важливий в випадках, коли важливо уникнути пропусків у виявленні позитивних класів [11].

Кожна з цих метрик має свою специфіку і використовується в залежності від того, на якій частині процесу класифікації потрібно сконцентрувати увагу. Наприклад, якщо важливо уникнути хибнопозитивних результатів, треба орієнтуватися на Precision, а якщо важливо захопити всі позитивні приклади, то потрібно підвищувати Recall. Ці метрики допомагають визначити різні аспекти продуктивності класифікаційних моделей й вибрати ту, яка найбільше відповідає вимогам завдання.

1.3 Методи первинної обробки даних

Методи первинної обробки даних [12] є важливим етапом в підготовці даних для подальшого аналізу та машинного навчання (рисунок 1.7). Очистка даних передбачає обробку відсутніх значень, що може включати їх вилучення чи заповнення, а також виявлення і виправлення помилкових значень, таких як аномалії чи викиди. Це допомагає уникнути спотворення результатів моделі.

Кодування категоріальних даних є важливим для перетворення змінних

в числовий формат. One-Hot Encoding перетворює категорії в бінарні ознаки, а Label Encoding присвоює числові значення кожному класу. Це робить дані більш зручними для алгоритмів машинного навчання.

Масштабування даних є критичним для нормалізації числових ознак. Нормалізація переводить дані в діапазон від 0 до 1, тоді як стандартизація забезпечує середнє значення 0 та стандартне відхилення 1, що важливо для деяких алгоритмів. Обробка даних, включаючи токенізацію та лематизацію, є важливими для аналізу даних. Токенізація розбиває текст на окремі одиниці, а лематизація допомагає привести слова до їх базових форм, що важливо для зменшення різноманітності словоформ.

Рисунок 1.7 - Методи первинної обробки даних

Вибір ознак за допомогою кореляційного аналізу дозволяє вибрати найбільш інформативні ознаки для моделі, а методи відбору ознак дозволяють автоматично визначити найбільш важливі характеристики для

класифікації або прогнозування. Обробка часових рядів включає згладжування даних для виявлення трендів та зменшення шуму. Експоненційне згладжування дозволяє зважувати більш нові дані, надаючи їм більше значення при прогнозуванні.

Обробка викидів передбачає використання статистичних методів для виявлення екстремальних значень та визначення, чи потрібно їх видаляти чи коригувати в залежності від контексту.

Розбиття даних є важливим етапом, зокрема стратифікація дозволяє зберегти пропорції класів при поділі на тренувальний та тестовий набори, а рандомізоване розбиття дозволяє забезпечити різноманітність вибірки.

Наведені методи допомагають гарантувати, що дані підходять для подальшого аналізу або моделювання, та що вони відповідають вимогам конкретного завдання [12]. Ретельна первинна обробка даних забезпечить більш точні та надійні результати в машинному навчанні, а також при аналізі даних.

Правильна підготовка даних, зокрема обробка пропущених значень, масштабування та вибір ознак, підвищує точність моделей. Метрики, такі як точність, чутливість та AUC, дозволяють оцінити ефективність моделей. Методи машинного навчання дають змогу ефективно обробляти складні дані та знаходити найкращі стратегії для прогнозування необхідності ТО. Вони дозволяють здійснювати попереджувальний аналіз для виявлення потенційних проблем та збоїв у роботі обладнання, що дає змогу запобігати непередбачуваним поломкам та зменшити витрати на аварійні ремонти.

Загалом, правильний підхід до аналізу даних та вибору відповідних методів прогнозування є ключовим для успішного управління технічним обслуговуванням, що дозволяє підвищити ефективність операцій та знизити витрати на підтримку обладнання в належному стані.

2 АРХІТЕКТУРА СИСТЕМИ ЗБОРУ ТА АНАЛІЗУ ДАНИХ ДЛЯ ПРОГНОЗУВАННЯ ТЕХНІЧНОГО ОБСЛУГОВУВАННЯ

2.1 Вибір технологій та платформи для реалізації системи

При виборі технологій та платформи для створення системи збору і аналізу даних для прогнозування ремонту та заміни обладнання важливо враховувати кілька аспектів. По-перше, обсяг даних — система повинна обробляти великі обсяги даних з різних джерел. По-друге, складність аналізу — необхідно мати можливість застосовувати складні методи, зокрема машинне навчання та штучний інтелект, для точного прогнозування. По-третє, важливо, щоб система була легко масштабованою та підтримуваною для адаптації до змін у вимогах та зростанні обсягу даних.

Для реалізації системи збору та аналізу даних для прогнозування ремонту та заміни обладнання можна виділити кілька ключових технологій (рисунок 2.1).

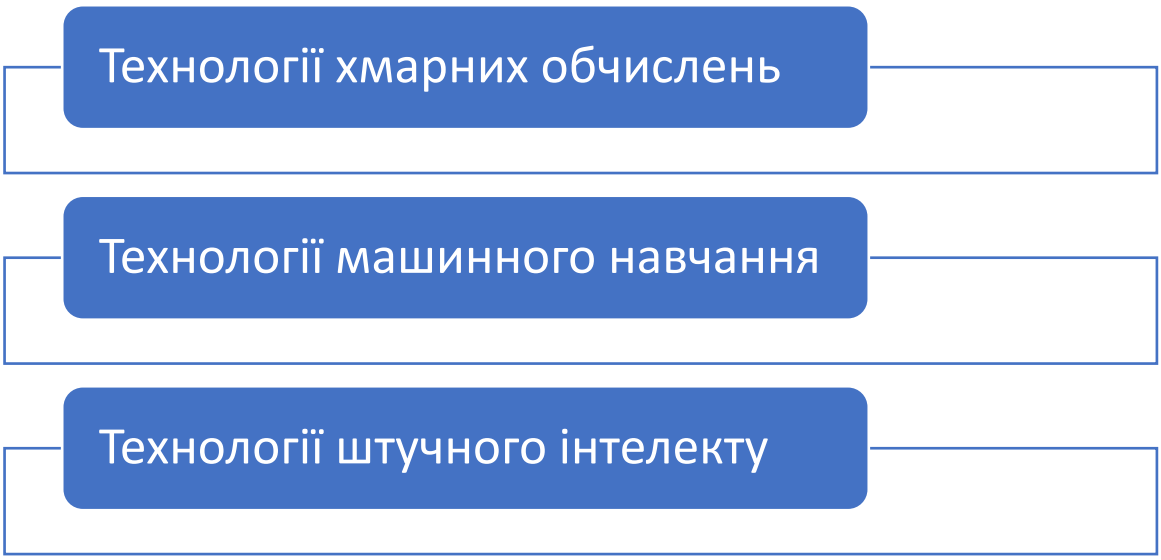


Рисунок 2.1 – Ключові технології реалізації системи збору та аналізу даних

По-перше, хмарні обчислення, які дозволяють легко масштабувати систему та забезпечують доступ до потужних ресурсів для обробки великих обсягів даних. По-друге, технології машинного навчання, що дають змогу проводити складний аналіз даних, що є необхідним для точного прогнозування. І, по-третє, технології штучного інтелекту, що автоматизують процеси аналізу, підвищуючи ефективність роботи системи.

Для розробки та тестування методів програмного моніторингу ТО на елеваторному комплексі використовується хмарна платформа Google Colab (рисунок 2.2). Вона є безкоштовною та публічною, й призначена для створення та виконання Jupyter Notebooks (JN) у хмарі [13]. JN – це веб-інтерфейс, який дозволяє комбінувати код, текст, візуалізації та інші мультимедійні елементи в одному документі. Цей інструмент оптимально підходить для наукових досліджень та аналізу даних, оскільки забезпечує інтерактивну роботу з великими наборами даних й ефективне тестування алгоритмів.

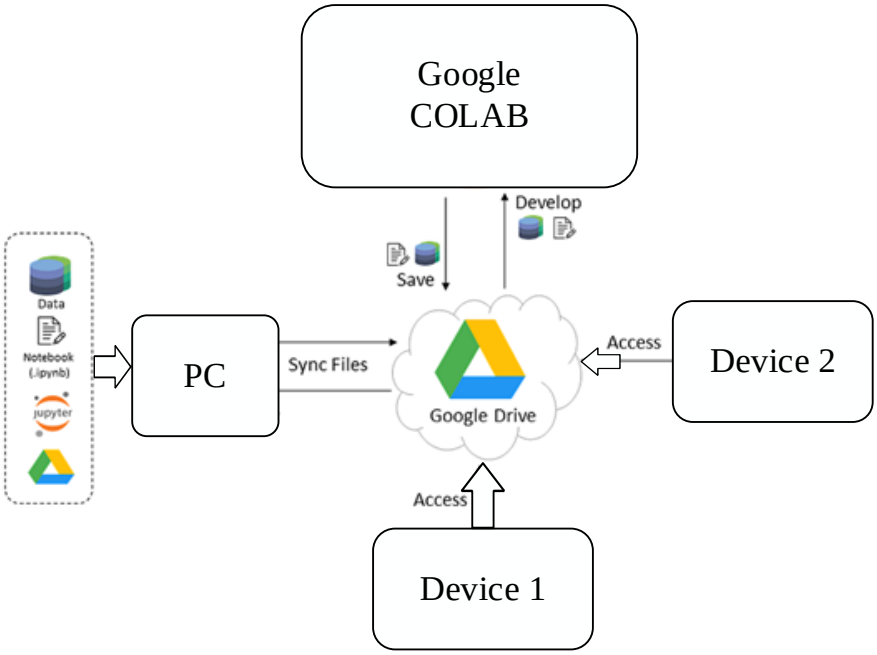


Рисунок 2.2 - Архітектура системи для прогнозування ТО
на базі платформи Google Colab

Google Colab є сучасною платформою для розробки та впровадження системи збору і аналізу даних для прогнозування ремонту та заміни обладнання. Платформа забезпечує використання декількох ключових технологій та інструментів, зокрема мови програмування Python, яка є основною для машинного навчання і штучного інтелекту. Для наукових обчислень використовується бібліотека NumPy, що дозволяє працювати з великими масивами даних. Для маніпулювання та аналізу даних зазвичай застосовується бібліотека Pandas, що надає зручні інструменти для очищення та обробки даних [13]. Для реалізації алгоритмів машинного навчання, таких як класифікація та регресія, використовується бібліотека Scikit-learn.

Google Colab дозволяє ефективно обробляти та аналізувати дані, а також застосовувати алгоритми машинного навчання для прогнозування ремонту та заміни обладнання [13]. Використання Google Colab для системи збору та аналізу даних має кілька переваг та недоліків. Однією з головних переваг є те, що Google Colab є безкоштовною платформою, що дозволяє знизити витрати на хмарні обчислення. Крім того, ця платформа є публічною, що дає можливість користувачам ділитися своїми результатами та кодом з іншими. Легка доступність через веб-браузер дозволяє користувачам працювати з платформою без необхідності встановлювати додаткове програмне забезпечення. З іншого боку, існують й недоліки. Один з них — обмежені ресурси, оскільки Google Colab надає лише обмежений доступ до обчислювальних потужностей на хмарній платформі Google Cloud Platform, що може бути недостатньо для великих або ресурсомістких завдань. Також варто звернути увагу на питання безпеки, оскільки Google Colab не гарантує таку ж високу безпеку, як платні платформи хмарних обчислень [13].

Технологія MySQL є важливим інструментом у сфері зберігання та обробки даних, забезпечуючи ефективне зберігання та масштабування великих обсягів інформації, що робить її незамінною для систем аналізу та прогнозування ремонту та обслуговування обладнання [14]. Однією з головних переваг MySQL є висока продуктивність, що дозволяє швидко

обробляти великі обсяги даних, що надходять з різних джерел, таких як датчики та системи моніторингу. Це критично важливо для систем, які виконують складні операції збору та аналізу даних.

Масштабованість MySQL є ще однією важливою характеристикою, яка дозволяє системам адаптуватися до зростаючих вимог, що виникають із розширенням діяльності компанії. Універсальність цієї СУБД також грає значну роль, оскільки вона забезпечує сумісність з різними платформами, що дозволяє інтегрувати її в різноманітні середовища та використовувати в різних областях.

Варто відзначити високий рівень безпеки MySQL, що гарантує захист конфіденційної інформації, що є важливим аспектом для систем, які обробляють важливі та чутливі дані, включаючи інформацію про технічний стан обладнання та історії обслуговування [14].

Для збору даних використовується система SCADA Citect, яка є потужною системою для збору даних з різноманітних пристроїв, завдяки широкій підтримці різних методів збору даних [15]. Вона підтримує численні протоколи передачі даних, зокрема Modbus, OPC UA, ODBC та інші, що дозволяє здійснювати ефективний та надійний обмін інформацією між різними типами пристроїв. Така сумісність забезпечує легкість інтеграції SCADA Citect з різними промисловими системами, дозволяючи забезпечити взаємодію з численними пристроями та технологіями.

Крім того, SCADA Citect має вбудовані драйвери для підключення деяких поширених пристроїв, таких як промислові контролери та датчики. Ці вбудовані драйвери спрощують процес підключення й інтеграції, роблячи систему більш універсальною та зручною для користувачів. Ще однією значущою особливістю є наявність API Citect [15], що дозволяє розробляти індивідуальні методи збору даних. Це надає можливість створювати спеціалізовані рішення для підключення додаткових пристроїв до системи, що значно підвищує гнучкість і адаптованість платформи до конкретних потреб користувачів. В результаті, SCADA Citect надає широкі можливості

для налаштування і розширення системи збору даних, що робить її надійним і універсальним інструментом для інтеграції в різноманітні промислові процеси. Використання SCADA Citect забезпечує розширені можливості для збору даних з різноманітних пристроїв, враховуючи різні протоколи, вбудовані драйвери та API для налаштування системи відповідно до конкретних вимог та характеристик промислових процесів.

2.2 Загальна архітектура системи

На основі вибраних технологій, таких як Google Colab, MySQL, SCADA Citect та інших, пропонується архітектура системи (рисунок 2.3), яка забезпечує збір, зберігання, аналіз даних та прогнозування ТО на елеваторному комплексі.

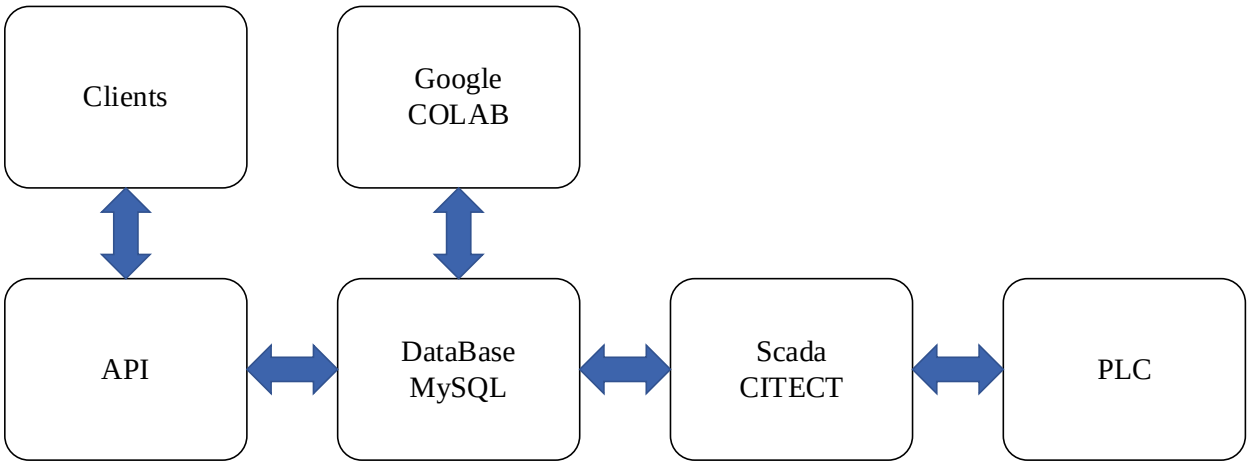


Рисунок 2.3 – Архітектура системи зберігання, аналізу даних та прогнозування ТО обладнання на елеваторному комплексі

Архітектура системи збору та аналізу даних для прогнозування ремонту та заміни обладнання складається з кількох ключових компонентів, які забезпечують ефективне функціонування та взаємодію системи,

спрямованої на оптимізацію ТО та управління промисловим обладнанням.

Програмований логічний контролер (PLC) є одним із центральних елементів системи. Він розроблений спеціально для роботи в промислових умовах, й автоматизує технологічні процеси. PLC виконує функції керування такими елементами, як конвеєрні лінії, насоси чи верстати з числовим програмним керуванням. У даній архітектурі PLC є джерелом даних, забезпечуючи збір важливих параметрів технологічного процесу, таких як температура, тиск, швидкість чи обсяги виробництва. Ці дані є основою для подальшого аналізу, прогнозування та прийняття управлінських рішень [16].

Для забезпечення диспетчерського управління та збору даних у режимі реального часу використовується система Supervisory Control And Data Acquisition (SCADA). Це програмне забезпечення виконує комплексні функції збору, обробки, відображення та архівування інформації, пов’язаної з моніторингом та управлінням промисловими процесами. SCADA дозволяє оперативно відстежувати стан обладнання, аналізувати ключові параметри й реагувати на відхилення у роботі систем.

API (інтерфейс програмування застосунків) виконує критично важливу роль у забезпеченні інтеграції між різними компонентами системи, зокрема між базою даних, інструментами аналізу та користувачем. Це інтерфейс, який визначає набір правил, протоколів і методів для взаємодії між програмами. API забезпечує стандартизовану та ефективну передачу даних, дозволяючи різним компонентам системи працювати разом у межах єдиного інформаційного середовища. Це значно спрощує інтеграцію системи, роблячи її більш гнучкою та масштабованою. Цей інтерфейс суттєво розширює можливості користувача у взаємодії з базою даних, забезпечуючи доступ до функцій отримання, оновлення та видалення даних без необхідності прямого звернення до базових таблиць або сховищ інформації. API виступає як посередник, який формує стандартизований формат обміну даними та забезпечує управління їх консистентністю, знижуючи ризики помилок й конфліктів.

Інтеграційна гнучкість API дозволяє легко об’єднувати різні системи та додатки, створюючи єдину екосистему програм. Це сприяє підвищенню взаємодії між ними, прискорюючи обмін інформацією й підвищуючи ефективність бізнес-процесів. У сучасному програмному середовищі, де швидкість та надійність передачі даних є критичними, API виконує роль ключового елемента, що підтримує розвиток систем і їхню адаптацію до нових технологій. Його використання забезпечує розширення можливостей програм і підвищує їх готовність до інтеграції з інноваційними сервісами.

Google Colab використовується як платформа для розробки та тестування методів аналізу даних, а також для навчання нейромережі на даних, що отримані від SCADA-системи. Ця платформа сприяє швидкому створенню, навчанню та вдосконаленню моделей машинного навчання, забезпечуючи зручний інструмент для роботи.

Зібрані SCADA-системою дані зберігаються у базі даних MySQL, яка підтримує різноманітні функції, зокрема масштабування та відновлення після збоїв. Ця СУБД забезпечує стабільне зберігання даних, необхідних для аналізу, що робить її невід'ємною частиною архітектури системи.

Для взаємодії з кінцевими користувачами використовується клієнтський інтерфейс. Він може бути реалізований як веб-додаток, мобільна програма або інше програмне забезпечення, яке забезпечує візуалізацію результатів аналізу даних. Це сприяє прийняттю ефективних рішень щодо ремонту або заміни обладнання, надаючи користувачам доступ до важливої інформації в зрозумілому вигляді.

Архітектура системи має низку важливих переваг. Завдяки гнучкості та масштабованості, SCADA-система може збирати дані з будь-якого промислового обладнання [15], а Google Cloud Platform забезпечує безпеку й адаптивність для зберігання даних та навчання моделей. Висока надійність цих компонентів гарантує стабільність роботи системи навіть у складних промислових умовах. Окрім того, архітектура залишається економічно вигідною завдяки використанню доступних технологій.

Проте система має й певні недоліки. Реалізація такого проєкту може бути складною через обсяги обладнання, що контролюється, та особливості використовуваних нейромереж. Крім того, значні витрати часу та ресурсів можуть знадобитися для навчання моделей, оскільки вони потребують великого обсягу якісних даних для досягнення високої точності прогнозів.

Запропонована архітектура для системи прогнозування ТО на елеваторному комплексі демонструє високу ефективність та гнучкість. Її адаптивність забезпечує можливість інтеграції з різноманітними виробничими процесами та масштабування відповідно до потреб підприємства. Використання SCADA-системи, бази даних MySQL та API дозволяє ефективно збирати, зберігати та обробляти інформацію з різних джерел, таких як датчики й системи моніторингу. Це створює основу для глибокого аналізу даних, необхідного для прогнозування технічного стану обладнання.

Завдяки платформі Google Colab та інструментам машинного навчання реалізується можливість розробки, тестування та впровадження алгоритмів, що забезпечують високу точність прогнозів [13]. Такий підхід сприяє підвищенню ефективності ТО, зниженню ризиків аварійних ситуацій та оптимізації витрат на ремонтні роботи. Водночас необхідність налаштування системи, навчання персоналу та обробки великих обсягів даних вимагають значних ресурсів, але ці інвестиції виправдані підвищенням надійності та стабільності роботи комплексу.

Таким чином, запропонована архітектура дозволяє створити інноваційну систему управління технічним обслуговуванням, яка поєднує сучасні технології збору, зберігання та аналізу даних, забезпечуючи стратегічну перевагу в управлінні елеваторним комплексом.

3 ВПРОВАДЖЕННЯ МЕТОДІВ ПРОГРАМНОГО МОНІТОРИНГУ ТЕХНІЧНОГО ОБСЛУГОВУВАННЯ У СИСТЕМУ ЗБОРУ ТА АНАЛІЗУ ДАНИХ НА ЕЛЕВАТОРНОМУ КОМПЛЕКСІ

3.1 Збір та підготовка даних

База даних є фундаментальною складовою будь-якої інформаційної системи, забезпечуючи збереження та доступ до інформації, необхідної для підтримки бізнес-процесів й прийняття рішень. Процес її проєктування передбачає створення структури, що відповідає вимогам й завданням.

Ключові принципи проєктування бази даних включають однозначність, яка гарантує унікальність кожного елемента даних; цілісність, що забезпечує правильність і узгодженість інформації; незалежність, яка дозволяє ізолювати зміни в одній частині бази даних від впливу на інші; а також надійність, що забезпечує стійкість до збоїв та можливість відновлення даних.

Процес проєктування бази даних охоплює декілька етапів. Спершу проводиться аналіз вимог, під час якого збираються та узагальнюються дані про потреби користувачів й бізнес-процеси. На основі цього створюється концептуальна модель, що визначає основні сутності та їх зв’язки. Далі концептуальна модель трансформується у логічну модель з урахуванням обраної системи управління базами даних. Після цього здійснюється фізичне проєктування, яке деталізує структуру таблиць, індексів і обмежень. На завершальному етапі база даних створюється за допомогою засобів СУБД, тестується й оптимізується.

Для забезпечення ефективності бази даних важливо дотримуватися певних правил. Це включає правильне визначення залежностей між таблицями, нормалізацію для уникнення дублювання даних, використання індексів для підвищення швидкості пошуку, встановлення зовнішніх ключів для підтримання зв’язків та обмежень, а також впровадження механізмів

безпеки для захисту даних.

Таким чином, проєктування бази даних є багатогранним процесом, який вимагає ретельного планування та врахування технічних і бізнесових аспектів. Від правильного виконання цього процесу залежить стабільність, ефективність та надійність функціонування інформаційної системи.

На основі аналізу вимог та архітектурних особливостей системи розробляється структура бази даних, яка забезпечує зберігання, організацію та доступ до даних, необхідних для прогнозування ремонту та заміни обладнання. Структура бази даних враховує всі ключові елементи системи, їхні взаємозв'язки та потреби у масштабованості й надійності.

На рисунку 3.1 представлена модель бази даних, яка демонструє основні таблиці, їх атрибути та зв'язки між ними. Ця модель є основою для фізичного впровадження бази даних й подальшої інтеграції з іншими компонентами системи.

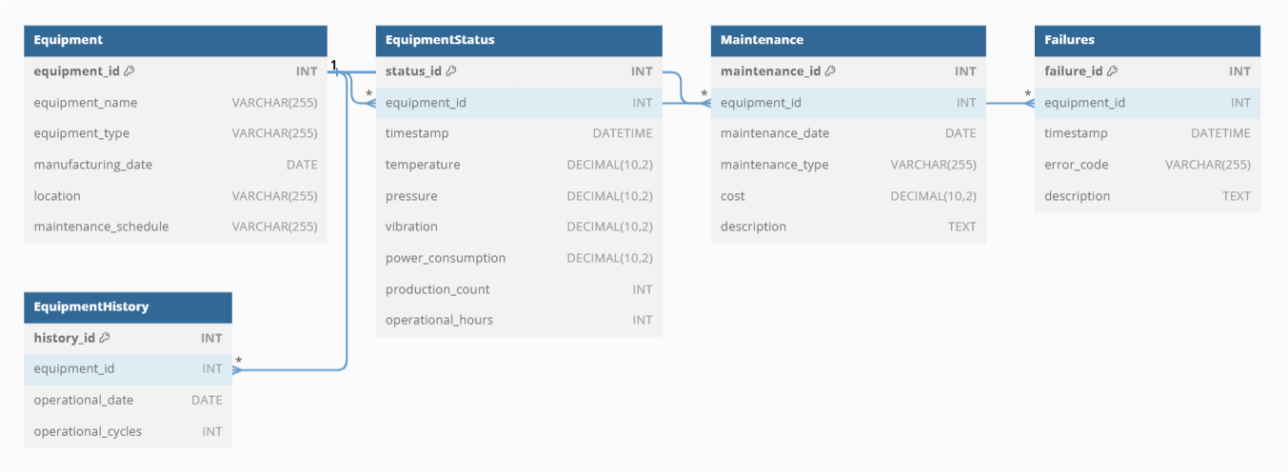


Рисунок 3.1 – Модель бази даних системи програмного моніторингу

Структура бази даних складається з кількох таблиць, кожна з яких відповідає за зберігання специфічної інформації, що використовується для аналізу, прогнозування ТО та управління обладнанням.

Таблиця Equipment (обладнання) є центральною у базі даних, оскільки

вона містить основну інформацію про обладнання. Унікальний ідентифікатор (equipment_id) забезпечує однозначне визначення кожної одиниці обладнання. Тут також зберігається назва, тип, дата виробництва, місцезнаходження та розклад обслуговування, що дозволяє впорядковувати та оптимізувати технічне обслуговування.

Таблиця EquipmentStatus (стан обладнання) зберігає дані про стан обладнання в реальному часі, такі як температура, тиск, вібрація, споживана потужність і кількість виробленої продукції. Це дозволяє оцінювати робочий стан обладнання та виявляти потенційні проблеми на ранніх стадіях. Поле timestamp фіксує час збору даних, що є важливим для аналізу трендів.

Таблиця Maintenance (обслуговування та ремонт) містить записи про обслуговування та ремонт обладнання. Вона включає інформацію про тип проведених робіт (планові чи аварійні), вартість, дату та детальний опис, що дозволяє відстежувати історію ТО кожного пристрою.

Таблиця Failures (відмови та несправності) зберігаються записи про відмови або несправності обладнання, включаючи час їх виникнення, коди помилок та детальний опис. Ця інформація дозволяє визначати типові проблеми та їх причини, що сприяє прогнозуванню відмов та покращенню планування ТО.

Таблиця EquipmentHistory (історія експлуатації) містить «історичні» дані про експлуатацію обладнання, включаючи кількість робочих циклів або інші оперативні показники за певні дати. Це допомагає оцінити продуктивність обладнання та визначати періоди найбільш інтенсивного навантаження.

Усі таблиці пов’язані через equipment_id, що забезпечує цілісність даних і дозволяє здійснювати комплексний аналіз усіх аспектів роботи обладнання. Наприклад, за допомогою зв’язків між таблицями можна відслідковувати, як певні умови експлуатації впливають на частоту відмов або витрати на ремонт.

Така структура забезпечує ефективну організацію даних та підтримує

процеси прогнозування ТО на основі зібраної інформації.

SCADA Citect є сучасним інструментом, що забезпечує диспетчерське управління та моніторинг виробничих процесів у реальному часі, пропонуючи ефективне рішення для автоматизації та контролю систем. Завдяки їй стає можливим не лише збір, але й раціональне використання даних, що сприяє оперативному реагуванню на зміни у виробничих процесах [15].

Інтерфейс головного екрана, представлений на рисунку 3.2, надає операторам деталізовану інформацію про стан системи. Він виконує функцію зв’язку між диспетчером й автоматизованою системою, забезпечуючи доступ до ключових показників роботи обладнання, стану сенсорів, параметрів процесів та інших важливих даних.

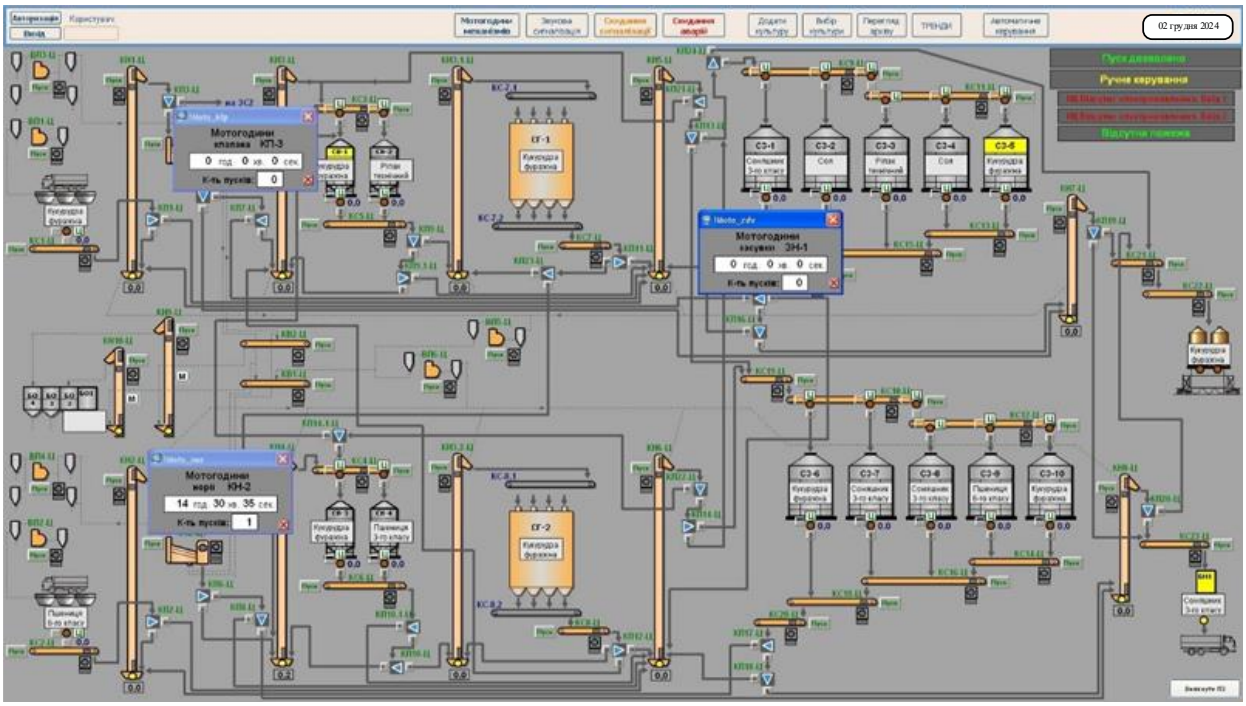


Рисунок 3.2 – Вікно відображення інформації щодо напрацювання обладнання елеваторного комплексу SCADA Citect

SCADA Citect забезпечує взаємодію з обладнанням, зокрема з PLC

контролерами, приводами та датчиками, завдяки підтримці різних протоколів зв'язку. Одним із таких протоколів є Modbus, що ілюструється на рисунку 3.3.

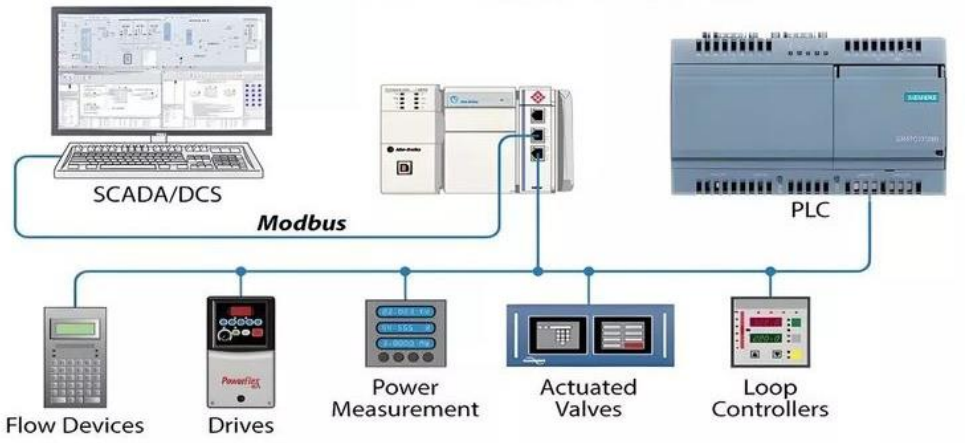


Рисунок 3.3 – Загальна схема мережі Modbus для керування технологічними процесами на елеваторному комплексі

Протокол Modbus являє собою промисловий стандарт обміну даними між електронними пристроями, що широко застосовується в автоматизації. Він забезпечує передачу інформації між програмованими логічними контролерами (PLC), сенсорами та іншими пристроями [17].

Процес інтеграції SCADA Citect із протоколом Modbus передбачає декілька етапів. Спочатку здійснюється конфігурація SCADA Citect, де створюються точки даних для зчитування nf запису інформації з обладнання. Далі налаштовується саме обладнання, яке буде взаємодіяти через Modbus, що включає в себе визначення адрес, швидкості передачі, типів даних та інших параметрів, необхідних для коректної роботи протоколу. Після цього встановлюється зв'язок між SCADA Citect та пристроями за допомогою відповідних драйверів або інструментів, що реалізують протокол Modbus. На наступному етапі система виконує зчитування/запис даних з пристроїв, що дає змогу отримувати інформацію про стан обладнання. Завершується процес

моніторингом та керуванням, під час якого SCADA Citect забезпечує візуалізацію зібраних даних, та дозволяє операторам контролювати процеси в реальному часі й оперативно реагувати на зміни. Такий процес забезпечує ефективну взаємодію SCADA Citect з обладнанням через протокол Modbus, дозволяючи здійснювати моніторинг та керування промисловими процесами. Запис даних, отриманих SCADA Citect, до бази даних MySQL може бути реалізований за допомогою стандартного інтерфейсу ODBC (Open Database Connectivity). Цей інтерфейс (рисунок 3.4) забезпечує взаємодію програм із різними СУБД, включаючи MySQL.

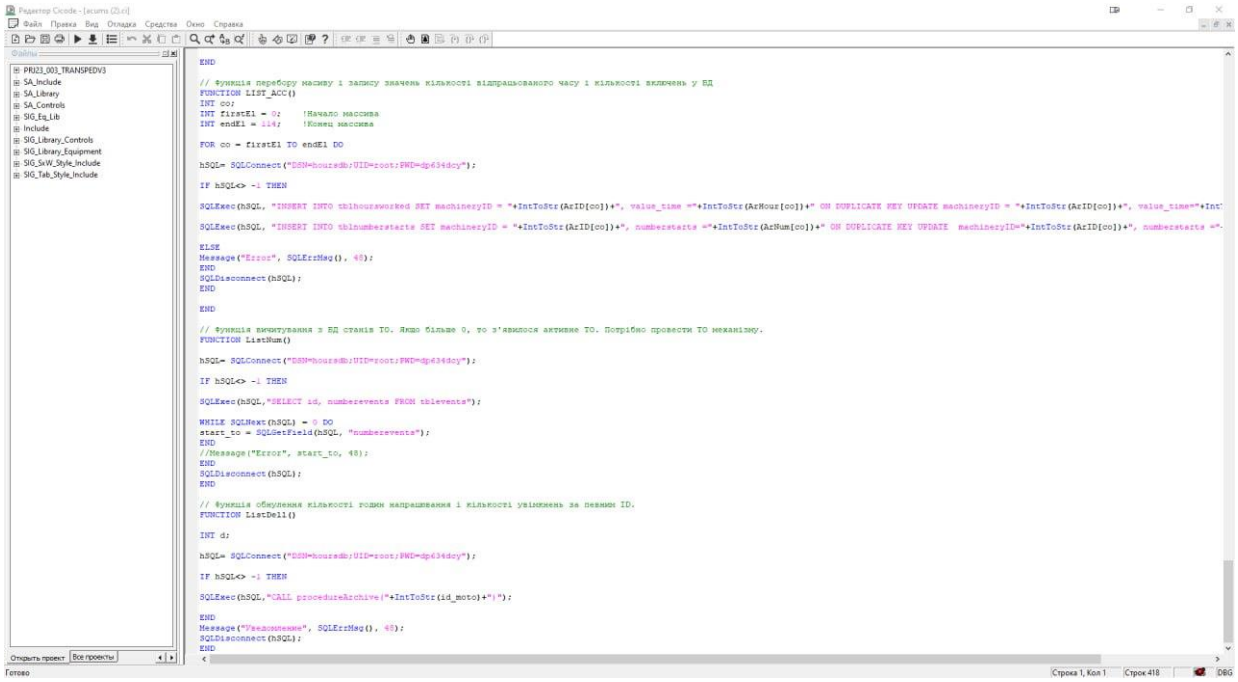


Рисунок 3.4 – Організація взаємодії SCADA Citect з БД MySQL

Процес запису даних у MySQL через ODBC у SCADA Citect передбачає кілька ключових етапів. Спершу виконується налаштування підключення, де визначаються параметри доступу до бази даних [18]. Далі формуються SQL-запити для запису або зчитування необхідної інформації, після чого ці запити інтегруються в конфігурацію SCADA-системи з використанням ODBC-функцій.

3.2 Розробка методів програмного моніторингу технічного обслуговування на елеваторному комплексі

Розробка методів програмного моніторингу ТО на елеваторному комплексі з використанням моделей машинного навчання відкриває нові можливості для автоматизації та оптимізації процесів управління обладнанням. Такий підхід дозволяє ефективно передбачати можливі поломки, здійснювати прогнозування залишкового ресурсу обладнання та здійснювати своєчасне технічне обслуговування, знижуючи витрати на ремонтні роботи та підвищуючи надійність систем.

Для передбачення залишкового ресурсу роботи обладнання широко застосовуються регресійні моделі, такі як лінійна регресія, логістична регресія та регресійні нейронні мережі. Аналізуючи дані про експлуатацію та телеметричні параметри, ці моделі дозволяють точно прогнозувати потребу в технічному обслуговуванні. Завдяки таким прогнозам можна визначати оптимальні моменти для проведення ремонту чи заміни обладнання, що сприяє безперебійному функціонуванню системи та підвищенню її продуктивності.

Моделі класифікації, такі як логістична регресія та дерева рішень, допомагають визначати ймовірність необхідності ТО у певний часовий проміжок. Це дозволяє підприємствам завчасно планувати роботи, мінімізувати втрати продуктивності та забезпечувати безперервну експлуатацію обладнання.

Моделі виявлення аномалій ефективно ідентифікують незвичні шаблони роботи, які можуть сигналізувати про потенційні несправності. Використання штучних нейронних мереж з рекурентними архітектурами, такими як LSTM (Long Short-Term Memory) та GRU (Gated Recurrent Unit), дозволяє виявляти відхилення в роботі обладнання та запобігати серйозним поломкам ще на ранніх етапах [12].

Успішне впровадження будь-якої моделі залежить від якості даних, які

використовуються для навчання. Інформація про історію експлуатації обладнання, графіки обслуговування та фактичні відмови є основою для ефективного аналізу. Ретельний збір та попередня обробка цих даних значною мірою впливають на точність прогнозів, роблячи систему передбачення технічного обслуговування важливим інструментом для підвищення ефективності підприємства.

На рисунку 3.5 показана логістична регресія, яка є одним із базових методів у машинному навчанні та статистиці, особливо в завданнях бінарної класифікації.

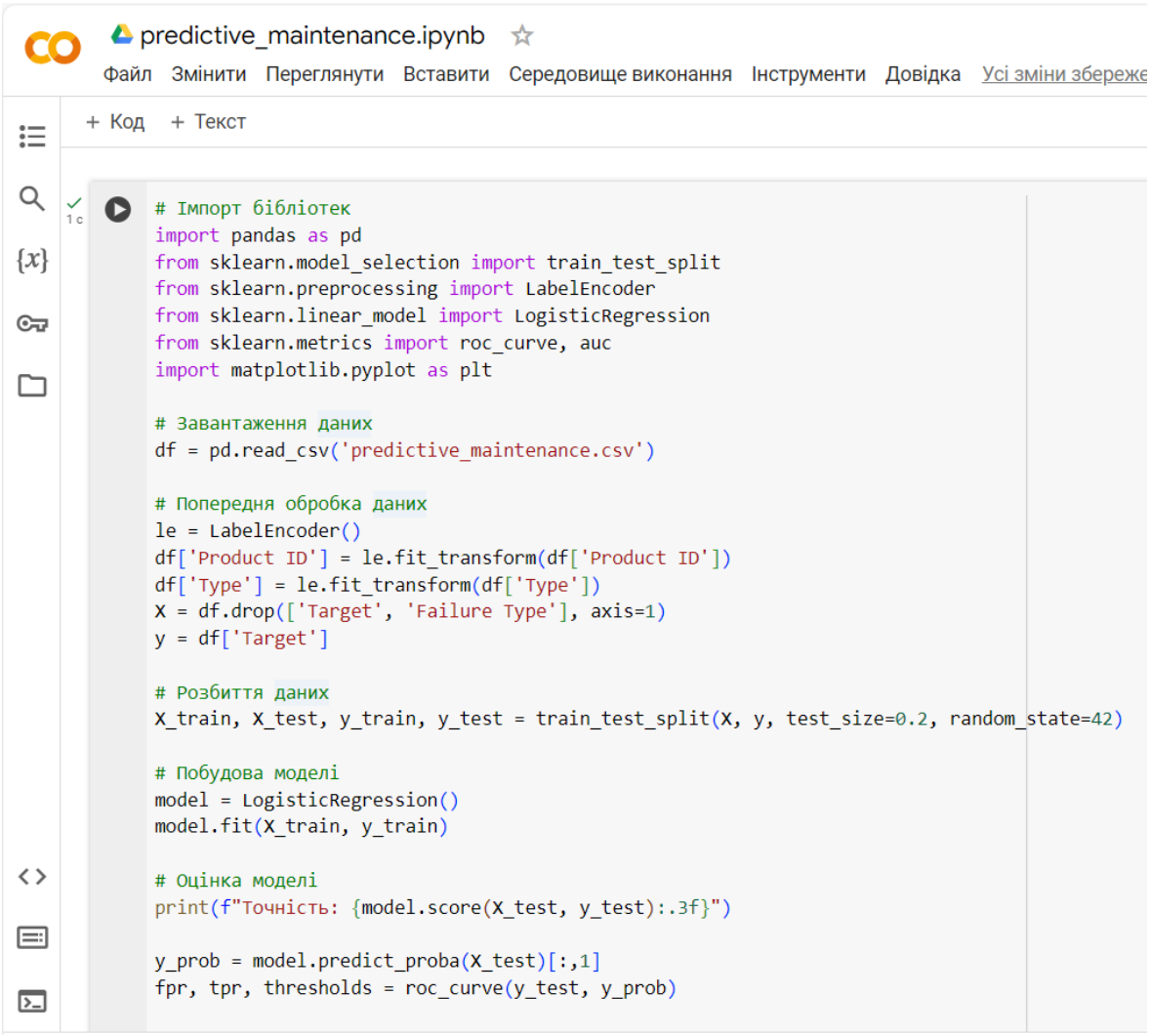


Рисунок 3.5 – Побудова моделі логістичної регресії для прогнозування відмов обладнання

Для її реалізації застосовується модуль `sklearn.linear_model` із бібліотеки Scikit-learn, що надає широкий спектр інструментів для побудови та аналізу моделей машинного навчання в Python. Scikit-learn вирізняється зручністю у використанні та високою ефективністю, що робить її незамінним ресурсом для аналітиків та розробників. Логістична регресія є однією з ключових технік для вирішення завдань класифікації завдяки своїй простоті, інтерпретованості та точності [18].

На рисунку 3.6 наведено використання моделі дерева рішень, реалізованої за допомогою модуля `sklearn.tree` із бібліотеки Scikit-learn.

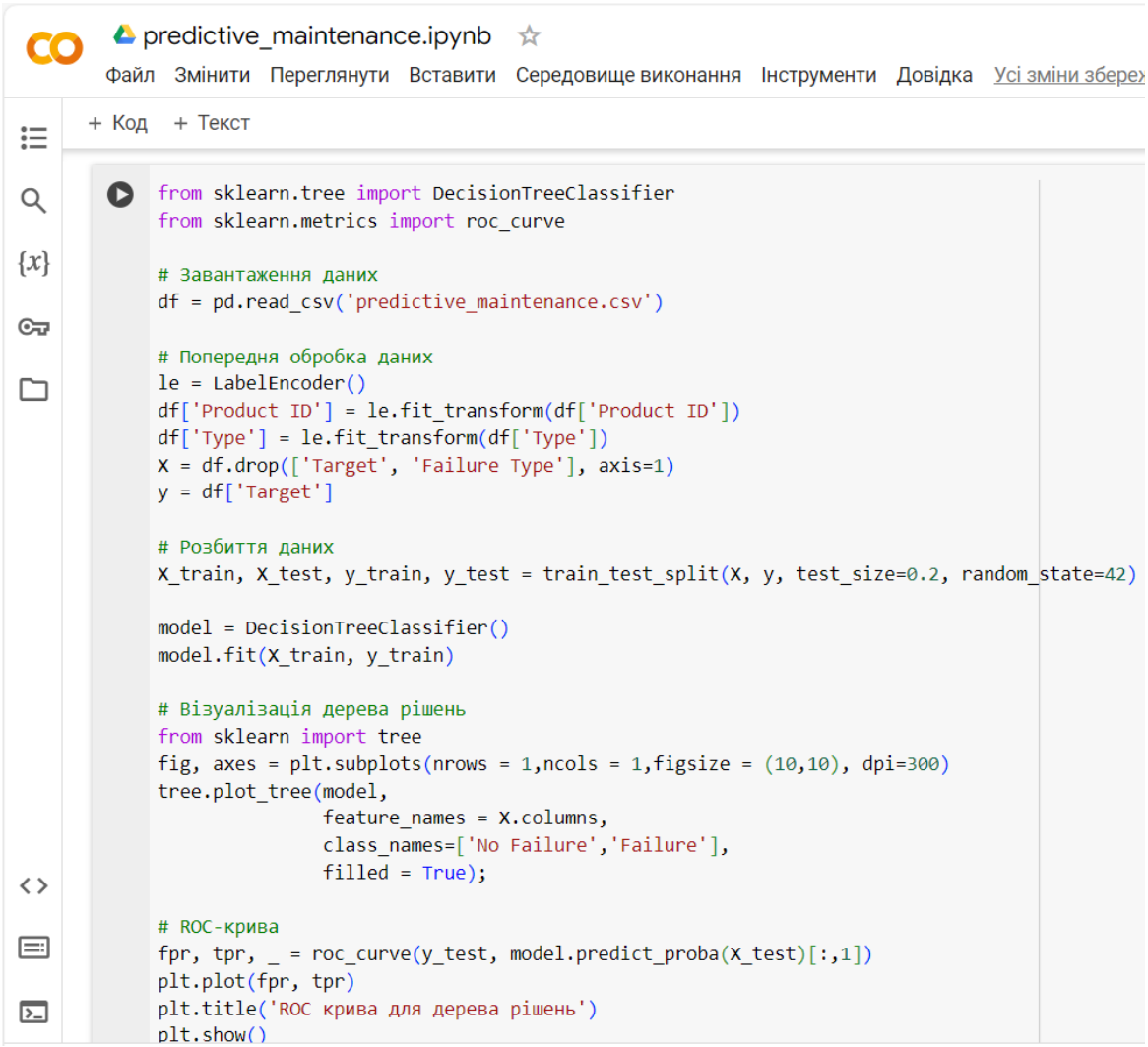


Рисунок 3.6 – Побудова моделі дерева рішень
для прогнозування відмов обладнання

Модуль забезпечує інтуїтивний інтерфейс для налаштування та впровадження дерев рішень, дозволяючи ефективно застосовувати методологію розгалужування та об'єднання для вирішення завдань машинного навчання. Scikit-learn, як потужний інструмент для роботи з даними у Python, забезпечує широкий спектр функціональних можливостей для створення та аналізу моделей.

Моделі дерев рішень є одним із ключових інструментів для класифікації та регресії, й вони дозволяють вирішувати складні задачі, пов’язані з прийняттям рішень на основі заданих умов і критеріїв. Крім того, дерева рішень сприяють виявленню важливих залежностей та структур у даних, що є важливим етапом їх аналізу та інтерпретації. Завдяки своїй гнучкості й наочності, цей метод широко використовується у вивченні та моделюванні складних систем.

На рисунку 3.7 показано побудову моделі випадкового лісу для прогнозування відмов обладнання. Для реалізації цієї моделі застосовується модуль `sklearn.ensemble` з бібліотеки Scikit-learn, який містить класи для роботи з ансамблевими методами, такими як випадковий ліс (Random Forest) та градієнтний бустінг (Gradient Boosting).

Ансамблеві методи, зокрема випадковий ліс, дозволяють комбінувати декілька дерев рішень для досягнення кращої точності прогнозів. Це значно підвищує стійкість моделі до випадкових варіацій даних і знижує ризик перенавчання, що робить їх ідеальними для вирішення складних завдань, таких як прогнозування відмов обладнання. Використання таких методів дозволяє отримати більш точні та надійні результати при аналізі великих і складних наборів даних.

На рисунку 3.8 показано побудову моделі SVM (Support Vector Machines) для прогнозування відмов обладнання. Для цього застосовується модуль `sklearn.svm` з бібліотеки Scikit-learn, що активно використовуються для завдань класифікації, регресії та виявлення аномалій у машинному навчанні [12].

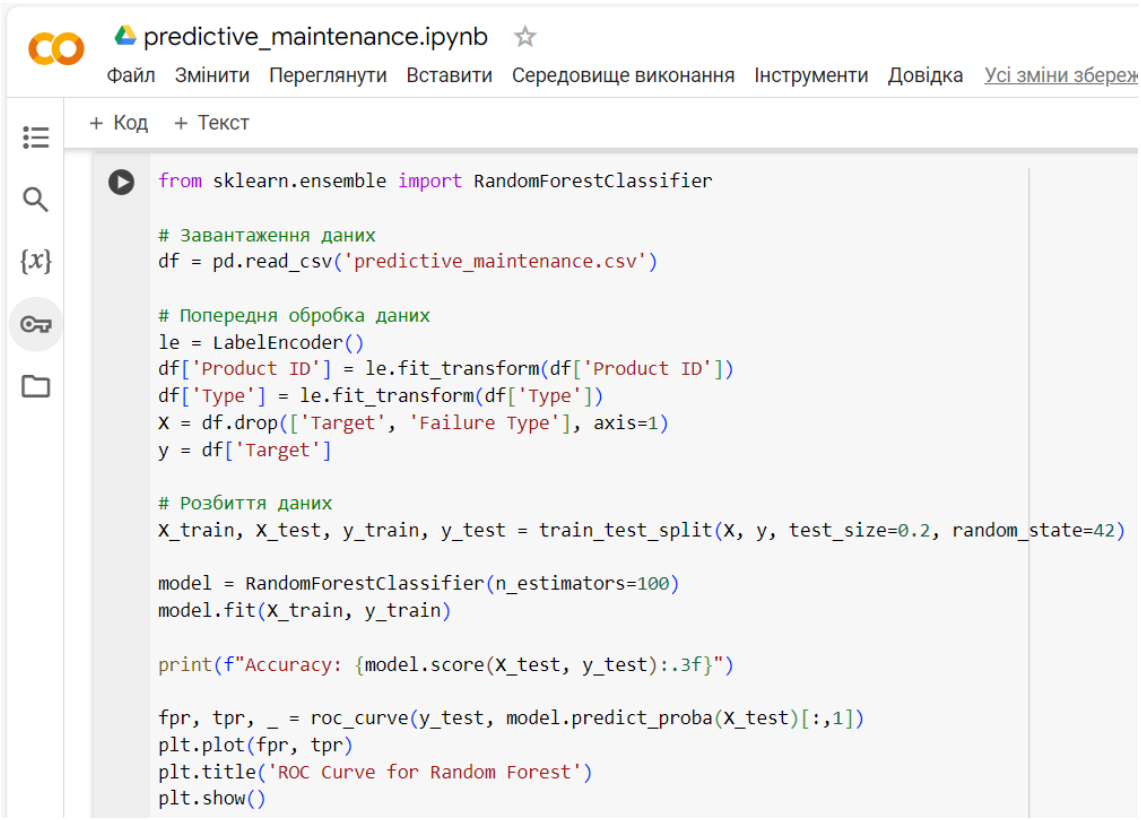


Рисунок 3.7 - Побудова моделі випадковий ліс
для прогнозування відмов обладнання

Метод опорних векторів є потужним інструментом для розв'язання складних задач, оскільки здатен ефективно розподіляти дані в багатовимірному просторі та знаходити оптимальні межі розділення між класами. Це дозволяє точно прогнозувати можливі відмови обладнання на основі аналізу історичних даних, що робить модель SVM корисною для моніторингу стану техніки та запобігання несправностей.

Створені методи на базі моделей машинного навчання демонструють значний потенціал у покращенні процесів прогнозування та ТО обладнання. Використання таких моделей, як регресія та класифікація, дозволяє ефективно прогнозувати необхідність ТО, оптимізуючи графіки ТО та знижуючи ризик непередбачених відмов. Це сприяє підвищенню продуктивності та зниженню витрат на неплановий ремонт. Методи ансамблевих моделей, зокрема випадковий ліс, дозволяють комбінувати

результати кількох дерев рішень, що значно покращує точність прогнозів і робить модель стійкішою до випадкових варіацій у даних. Це дає змогу враховувати численні фактори, що можуть впливати на працездатність обладнання, що важливо для забезпечення його безперебійної роботи.

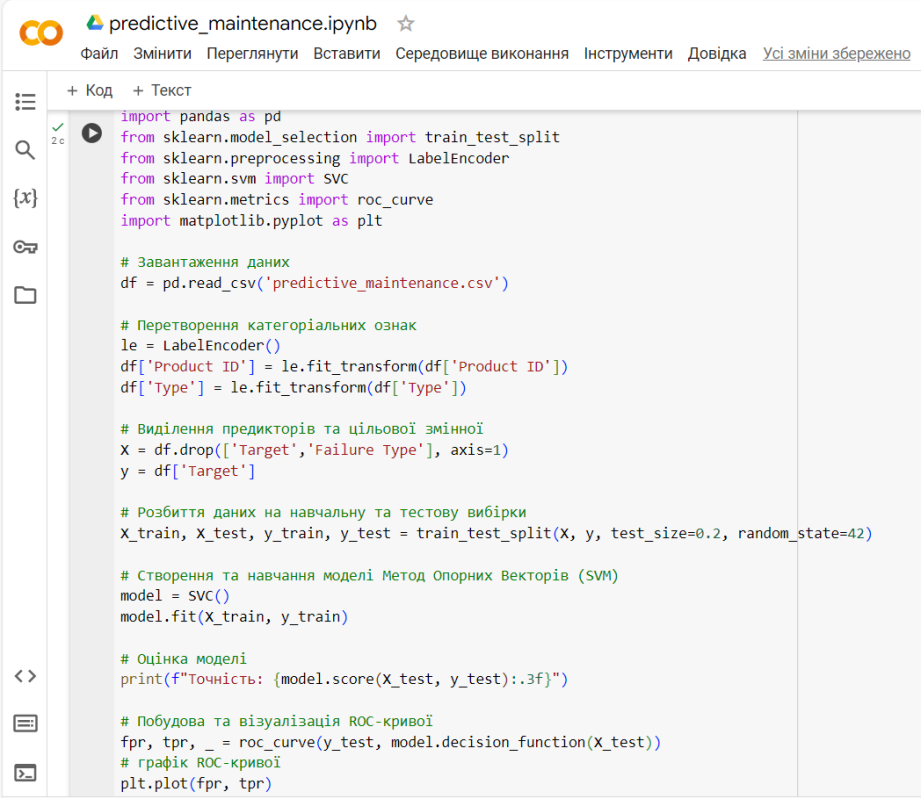


Рисунок 3.8 - Побудова моделі SVM для прогнозування відмов обладнання

Методи класифікації, такі як логістична регресія та дерева рішень, допомагають визначити ймовірність необхідності ТО в конкретний період, що дозволяє заздалегідь планувати роботи, уникати термінових втрат продуктивності та забезпечувати ефективну роботу обладнання. Загалом, застосування цих методів дозволяє значно підвищити ефективність та надійність систем моніторингу та обслуговування обладнання, забезпечуючи стабільну роботу та зниження витрат у довгостроковій перспективі.

4 ДОСЛІДЖЕННЯ МЕТОДІВ ПРОГРАМНОГО МОНІТОРИНГУ

Проведення досліджень методів програмного моніторингу ТО на елеваторному комплексі є важливим кроком у забезпеченні безперебійної та ефективної роботи обладнання. Такі дослідження спрямовані на впровадження інноваційних рішень для прогнозування, планування та контролю ТО, що дозволяє знизити витрати, мінімізувати простої та підвищити продуктивність комплексу.

4.1 Опис експериментальних даних

Оскільки реальні набори даних про прогнозоване технічне обслуговування зазвичай є недоступними через їх конфіденційність або складність збору, у дослідженні використано синтетичний набір даних, який імітує характерні параметри та шаблони, що спостерігаються в галузі. Такий підхід дозволяє моделювати процеси прогнозування ТО, використовуючи дані, максимально наближені до реальних умов. На рисунку 4.1 наведено приклад цього набору даних, який відображає ключові характеристики, такі як параметри роботи обладнання, індикатори стану та часові інтервали між обслуговуваннями.

Синтетичні дані створюють можливість тестування й оптимізації методів прогнозування, дозволяючи аналізувати їхню ефективність без ризику для реального обладнання. Це забезпечує основу для перевірки моделей та алгоритмів машинного навчання, а також оцінки їх здатності точно визначати час і характер необхідного ТО.

Синтетичний набір даних, використаний у дослідженні, складається зі 10000 точок даних, представлених у форматі таблиці. Кожен рядок цієї таблиці містить 14 характеристик, що відображають різноманітні параметри стану обладнання та умови його роботи. Характеристики включають такі

показники, як робоча температура, вібрація, навантаження, рівень зношення, кількість годин роботи після останнього обслуговування, а також інші ключові параметри, що впливають на технічний стан обладнання.

model1.py | Preview 'predictive_maintenance.csv'

	UDI	Product ID	Type	Air temperatur	Process temper	Rotational spee	Torque [Nm]	Tool wear [min]	Target	Failure Type
	1	M14860	M	298.1	308.6	1551	42.8	0	0	No Failure
	2	L47181	L	298.2	308.7	1408	46.3	3	0	No Failure
	3	L47182	L	298.1	308.5	1498	49.4	5	0	No Failure
	4	L47183	L	298.2	308.6	1433	39.5	7	0	No Failure
	5	L47184	L	298.2	308.7	1408	40	9	0	No Failure
	6	M14865	M	298.1	308.6	1425	41.9	11	0	No Failure
	7	L47186	L	298.1	308.6	1558	42.4	14	0	No Failure
	8	L47187	L	298.1	308.6	1527	40.2	16	0	No Failure
	9	M14868	M	298.3	308.7	1667	28.6	18	0	No Failure
	10	M14869	M	298.5	309	1741	28	21	0	No Failure
	11	H29424	H	298.4	308.9	1782	23.9	24	0	No Failure
	12	H29425	H	298.6	309.1	1423	44.3	29	0	No Failure
	13	M14872	M	298.6	309.1	1339	51.1	34	0	No Failure
	14	M14873	M	298.6	309.2	1742	30	37	0	No Failure
	15	L47194	L	298.6	309.2	2035	19.6	40	0	No Failure
	16	L47195	L	298.6	309.2	1542	48.4	42	0	No Failure
	17	M14876	M	298.6	309.2	1311	46.6	44	0	No Failure
	18	M14877	M	298.7	309.2	1410	45.6	47	0	No Failure
	19	H29432	H	298.8	309.2	1306	54.5	50	0	No Failure
	20	M14879	M	298.9	309.3	1632	32.5	55	0	No Failure

Рисунок 4.1 – Фрагмент набору даних класифікації предиктивного ТО

Такий обсяг та структура даних дозволяють забезпечити всебічний аналіз, необхідний для розробки моделей прогнозування ТО. Наявність широкого спектра характеристик у кожній точці даних забезпечує можливість побудови точних моделей машинного навчання для прогнозування часу обслуговування, виявлення аномалій та оцінки ризику відмов [7].

Синтетичний набір даних, створений для дослідження прогнозування ТО, має чітку структуру й включає всі необхідні параметри, що відображають реальні умови роботи обладнання. Унікальний ідентифікатор для кожного запису представлений значенням UID, яке забезпечує однозначну ідентифікацію кожної точки даних у межах набору. Також важливою складовою є поле productID, яке відображає якість продукції за

трьома категоріями: низька, середня та висока. Ці категорії позначаються відповідними літерами L, M та H, де 50% даних відповідають низькій якості, 30% — середній, а 20% — високій. Додатково productID містить серійний номер, що дає змогу деталізувати окремі варіанти продукції.

Температура повітря моделюється за допомогою методу випадкового блукання, після чого нормалізується з урахуванням стандартного відхилення 2 К у межах діапазону близько 300 К. Температура процесу, яка є невід'ємною складовою технологічного середовища, базується на температурі повітря з додаванням постійної величини 10 К і корекції за допомогою випадкового блукання, нормалізованого до стандартного відхилення 1 К.

Швидкість обертання обладнання представлена у вигляді значень, розрахованих на основі потужності 2860 Вт. Для забезпечення реалістичності ці значення доповнені нормально розподіленим шумом. Крутний момент, що є ще одним ключовим параметром, має значення, розподілені навколо середнього значення 40 Н·м із стандартним відхиленням 10 Н·м, при цьому жодне значення не може бути від'ємним.

Окрему увагу приділено характеристиці зносу інструменту, що залежить від якості продукції. Продукти високої, середньої та низької якості додають до зносу інструменту відповідно 5, 3 і 2 хвилини, що відображає вплив якості продукції на ресурс обладнання. Крім того, включена інформація про відмови верстата, яка зазначає, чи стався збій у конкретній точці даних. Вказані режими відмов дозволяють враховувати потенційні ризики та моделювати відповідні заходи для їх попередження.

Таким чином, цей набір даних є багатогранним інструментом для аналізу та моделювання, й дозволяє враховувати всі ключові аспекти, які впливають на стан та продуктивність обладнання. Завдяки своїй структурі він є корисним для імітації реальних виробничих сценаріїв, що сприяє розробці та впровадженню ефективних методів прогнозування ТО.

Гістограма розподілу ознак є одним із ключових інструментів для візуалізації даних, що дозволяє зрозуміти, як значення конкретної ознаки розподілені в межах вибірки. На вертикальній осі гістограми представлено частоту або відсоткову частку входжень значень, а на горизонтальній осі — діапазон значень самої ознаки. Такий підхід дає змогу швидко оцінити властивості розподілу, включаючи його форму, симетрію, наявність викидів або аномалій, а також можливі кластеризації.

У контексті машинного навчання гістограми є важливими для виконання кількох завдань. По-перше, вони сприяють розумінню розподілу даних, що є основою для вибору підходящих алгоритмів моделювання. Наприклад, моделі, які припускають нормальний розподіл даних, вимагають перевірки та, за необхідності, трансформації ознак. По-друге, гістограми дозволяють виявити викиди, які можуть спотворювати результати моделювання та потребують особливого підходу. По-третє, ці графічні представлення допомагають визначити важливі ознаки для моделювання, оскільки рівномірність або вузький діапазон значень можуть свідчити про їхню низьку інформативність. Нарешті, вони полегшують вибір параметрів моделі, таких як кількість бінів для категоризації даних або масштаби для нормалізації.

Для створення гістограми розподілу ознак синтетичного набору даних було використано модуль seaborn із бібліотеки Python. Цей модуль забезпечує високу якість візуалізації та широкі можливості налаштування графіків, що дозволяє деталізувати аналіз кожної ознаки. На рисунку 4.2 представлено результати аналізу, які підтверджують важливість розуміння розподілу ознак для побудови ефективних моделей машинного навчання.

Аналіз гістограм дає змогу оцінити розподіл основних ознак синтетичного набору даних, що має ключове значення для розуміння їхніх властивостей перед використанням у машинному навчанні.

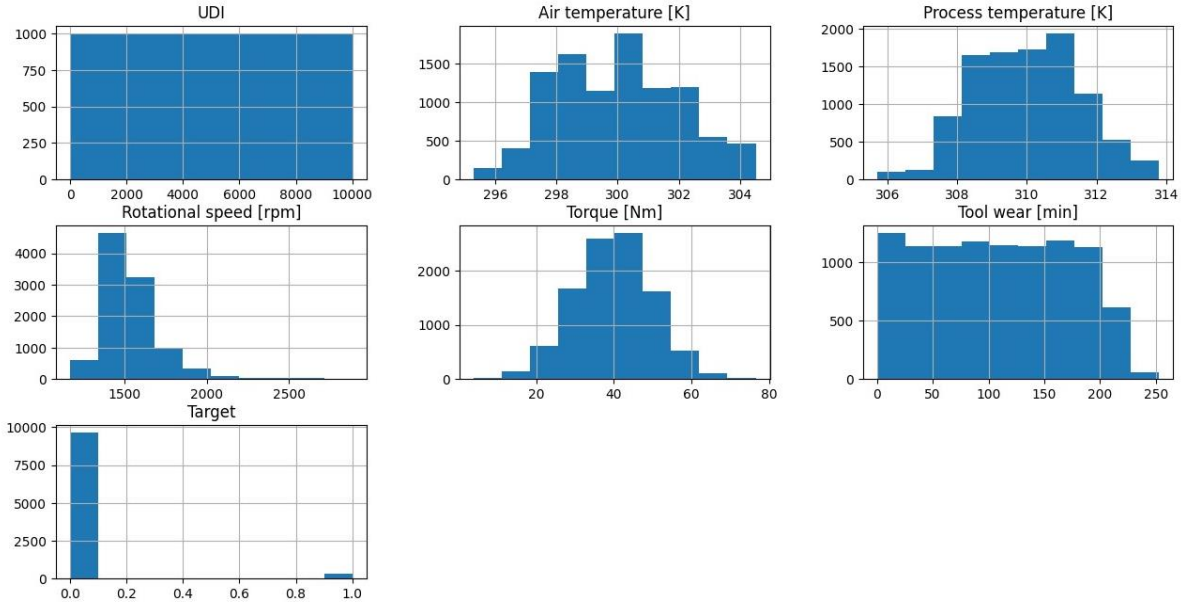


Рисунок 4.2 - Гістограми розподілів ознак

Середнє значення температури повітря (Air temperature) становить близько 1250 К, що вказує на центральне положення цього показника в наборі даних. Значна кількість значень знаходиться в межах від 1000 до 1500 К, що демонструє типове робоче середовище. Температури рідко опускаються нижче 500 К або перевищують 2000 К, що може бути наслідком обмежень у процесі або специфічного налаштування синтетичного набору даних.

Середня температура процесу (Process temperature) становить близько 1500 К, що вище за середню температуру повітря, що є очікуваним для промислових процесів. Найчастіше температура процесу коливається в межах 1250–1750 К. Винятки, нижчі за 1000 К або вищі за 2000 К, є малоймовірними, що свідчить про стабільність контролю цього показника в синтетичному моделюванні.

Середній крутний момент (Torque) становить 5000 Нм. Його значення найчастіше знаходяться в межах 2500–7500 Нм, що відображає типову експлуатаційну область. Занадто низькі (нижче 0 Нм) або високі (понад 10000 Нм) значення практично не зустрічаються, що відповідає фізичним

обмеженням і реалістичності синтетичних даних.

Середній знос інструменту (Tool wear) дорівнює приблизно 10000 хв, що вказує на типовий термін експлуатації. Найбільше значень зосереджено в діапазоні 5000–15000 хв. Виняткові значення, нижчі за 0 хв або вищі за 20000 хв, є рідкісними й можуть вказувати на обмеження терміну служби інструменту.

Середня швидкість обертання (Rotational speed) становить близько 4000 об/хв, а більшість значень розташовані між 2000 і 6000 об/хв. Значення нижче 0 об/хв або вище 8000 об/хв є малоймовірними, що може бути наслідком фізичних обмежень обладнання або обмежень у синтетичному наборі даних.

Середнє значення цільової змінної (Target) становить близько 150. Найбільше значень зосереджено в межах 50–250. Надзвичайно низькі (менше 20) або високі (понад 250) значення цілі є рідкісними, що свідчить про добре збалансований набір даних.

Такий аналіз показує, що синтетичний набір даних має стабільні та реалістичні характеристики. Його особливості добре відповідають типовим робочим умовам, що дозволяє ефективно використовувати ці дані для навчання та тестування моделей машинного навчання.

Загальний аналіз гістограм демонструє, що процеси в системі характеризуються стабільністю параметрів, попри їх високі робочі значення. Температура повітря та процесу показують чітку тенденцію до нагріву, утримуючись у визначених межах. Це свідчить про контрольованість умов роботи, що є критично важливим для забезпечення безпеки та ефективності функціонування обладнання.

Крутний момент, маючи високі значення, варіюється в залежності від зносу інструменту. Така залежність може бути використана для прогнозування терміну служби інструменту та оптимізації його заміни, що є одним із ключових аспектів забезпечення безперервної роботи процесу. Швидкість обертання також демонструє стабільно високі показники, що

можуть адаптуватися залежно від поставлених цілей, дозволяючи системі гнучко реагувати на зміну умов чи потреб.

Надані дані можна успішно використовувати для оцінки ефективності процесу. Наприклад, перевищення допустимих температурних меж може бути сигналом про потенційні ризики, такі як перегрів або пошкодження обладнання. Аналогічно, низький рівень зносу інструменту або відхилення від типових значень швидкості обертання можуть вказувати на неефективне використання ресурсів або зниження продуктивності.

Таким чином, аналіз цих параметрів є основою для створення ефективних моделей програмного моніторингу. Вони можуть не лише виявляти потенційні проблеми, але й пропонувати рекомендації для їх вирішення, що сприяє зниженню ризиків, підвищенню продуктивності та продовженню терміну експлуатації обладнання.

4.2 Оцінка ефективності

Оцінка ефективності моделей прогнозування є важливим етапом для забезпечення надійності прогнозів та впровадження моделей у реальні процеси. Аналіз результатів моделі логістичної регресії, побудованої на змодельованих даних, показує, що цей метод демонструє високу точність у визначенні ймовірностей поломок.

Логістична регресія дозволяє ефективно класифікувати об'єкти та оцінювати ризики виникнення несправностей за допомогою аналізу взаємозв'язків між різними характеристиками. Отримані результати підтверджують, що модель здатна надавати обґрунтовані прогнози, дозволяючи підприємствам краще планувати технічне обслуговування та знижувати ризики несподіваних збоїв.

Одним із ключових аспектів роботи моделі є її здатність інтерпретувати вплив кожної характеристики на ймовірність поломки, що робить логістичну регресію зручною для впровадження в системи, де

важливий як точний прогноз, так і розуміння причин можливих відмов. Це відкриває можливості для адаптації процесів, спрямованих на мінімізацію впливу факторів ризику, що в підсумку сприяє підвищенню ефективності експлуатації обладнання.

На рисунку 4.3 наведена ROC-крива (Receiver Operating Characteristic curve), яка є важливим інструментом для оцінки ефективності класифікаційних моделей, зокрема у задачах бінарної класифікації, таких як прогнозування поломок обладнання. ROC-крива відображає залежність між чутливістю (True Positive Rate, TPR) та специфічністю (False Positive Rate, FPR) для різних порогів класифікації. Чутливість показує частку правильних позитивних прогнозів, тоді як специфічність відображає частку помилково класифікованих негативних випадків.

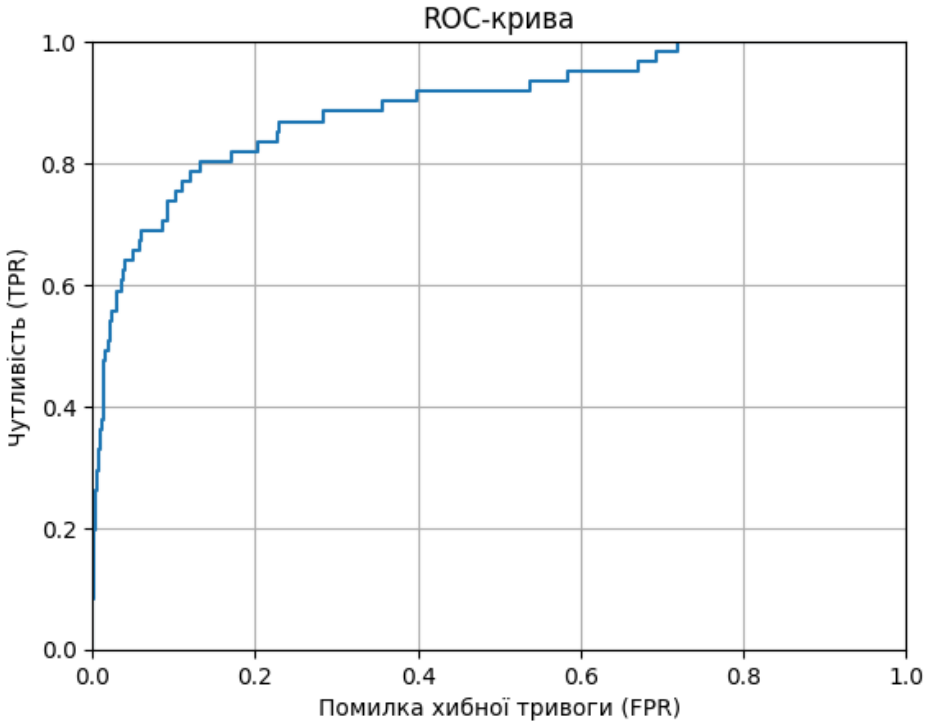


Рисунок 4.3 – ROC (Receiver Operating Characteristic) - крива

Отримані результати оцінки моделі логістичної регресії демонструють її високу ефективність у прогнозуванні ймовірностей поломок. Точність

моделі на рівні 0,974 свідчить про її здатність правильно класифікувати майже 97,4% випадків, що є дуже хорошим результатом для більшості практичних застосувань. Така точність свідчить про рідкісні помилки в класифікації та відносно невисокий рівень хибних спрацьовувань моделі.

AUC значення 0,895 також є сильним показником, оскільки наближається до максимально можливого значення 1. Це вказує на те, що модель добре справляється з розрізненням між позитивними і негативними класами (наприклад, поломками та відсутністю поломки). Чим ближче AUC до 1, тим точніше модель класифікує випадки.

Загалом, ці показники підтверджують, що модель логістичної регресії є надійним інструментом для прогнозування поломок обладнання. Вона може бути застосована для раннього виявлення потенційних проблем і дозволить підприємствам вчасно вжити заходів для запобігання серйозним поломкам та зниження витрат на ТО.

Модель логістичної регресії може бути ефективно використана для виявлення поломок, базуючись на даних, які регулярно збираються з різних джерел, таких як датчики, показники експлуатації та інформація про технічне обслуговування. Ці дані дозволяють моделі проводити точні прогнози щодо ймовірності виникнення поломок й можуть бути використані для оптимізації графіків ТО.

Один із основних напрямків використання моделі полягає в розробці планів ТО, які допоможуть запобігти непередбаченим поломкам. Це дозволяє підприємствам не лише знизити ризики поломок, але й скоротити витрати на аварійні ремонти, збільшити ефективність і довговічність обладнання. Крім того, модель може використовуватись для оцінки ефективності заходів, спрямованих на запобігання поломок. Це дозволяє виявити, чи досягнуті бажані результати завдяки впровадженим стратегіям ТО.

Попри високу точність, важливо зазначити, що модель логістичної регресії не є безпомилковою. Вона може робити помилки, особливо якщо надані дані для навчання моделі є неповними, нечіткими або неякісними.

Тому для досягнення найкращих результатів необхідно забезпечити наявність точних, повних і репрезентативних даних. Незважаючи на це, модель логістичної регресії є потужним і корисним інструментом, який допомагає значно підвищити надійність обладнання та оптимізувати процеси ТО на підприємстві.

Модель дерева рішень, на основі змодельованих даних (рисунок 4.4), продемонструвала високу точність у прогнозуванні ймовірностей поломок.

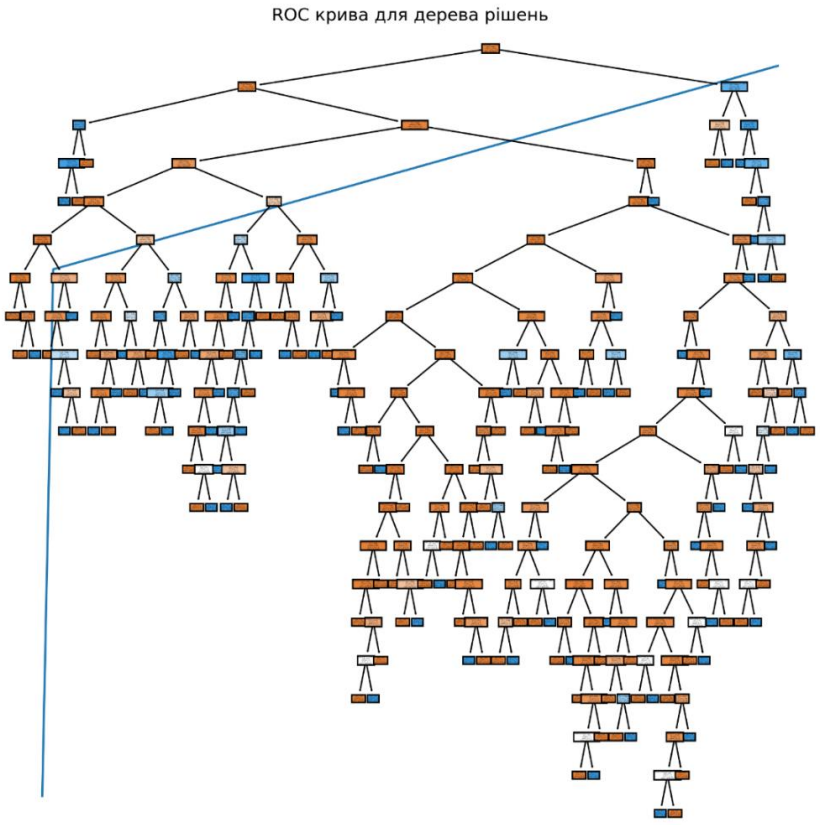


Рисунок 4.4 – ROC крива для дерева рішення

Вона є ефективною завдяки своїй здатності розбивати складні проблеми на простіші, що дозволяє чітко визначити, які саме умови можуть призвести до відмови обладнання. Дерево рішень будує послідовність правил, які допомагають зрозуміти, за яких обставин ймовірність поломки є найбільшою, а також виявляє ключові фактори, які мають найбільший вплив на це.

Ця модель дозволяє зрозуміти, як різні характеристики: температура, знос інструменту, швидкість обертання тощо, взаємодіють між собою й як вони впливають на ймовірність виникнення поломки. Також, дерево рішень дозволяє легко інтерпретувати результати, що є важливим для прийняття рішень на основі прогнози.

Дерево рішень може бути застосоване для виявлення ключових факторів, які потребують уваги при проведенні ТО, а також для формування планів профілактичного ремонту, знижуючи ймовірність серйозних поломок. Це дозволяє підприємствам не тільки знизити витрати на ремонт, але й підвищити загальну ефективність роботи обладнання.

Однак важливо відзначити, що дерево рішень може бути чутливим до перевищення глибини дерева, що може призвести до перенавчання. Це означає, що надмірно складні моделі можуть погано працювати на нових, невідомих даних.

Точність моделі дерева рішень становить 0,977, що вказує на її високу ефективність в класифікації та прогнозуванні ймовірностей поломок. Це означає, що модель правильно класифікує 97,7 % випадків, що є дуже хорошим результатом для задач прогнозування.

ROC-крива моделі також вказує на її високий рівень ефективності, оскільки вона демонструє хорошу здатність правильно відрізняти позитивні випадки (поломки) від негативних (відсутність поломок). Це свідчить про те, що модель може точно оцінювати ймовірність поломок навіть при різних порогах.

З огляду на ці показники, можна зробити висновок, що модель дерева рішень є надійним інструментом для прогнозування ймовірностей поломок на підприємствах, зокрема в системах ТО обладнання. Вона може допомогти виявити потенційні поломки на ранніх етапах, що дозволяє здійснити своєчасне технічне обслуговування й запобігти серйозним поломкам, які могли б призвести до значних витрат або зупинки виробничих процесів.

Модель дерева рішень є надзвичайно корисним інструментом для

прогнозування поломок, оскільки вона здатна працювати з різноманітними даними, що надходять із різних джерел, таких як датчики, дані з експлуатації та ТО. Вона дозволяє не лише визначити ймовірність майбутніх поломок, але й допомагає в оптимізації процесів ТО. Таке прогнозування є важливим для розробки планів ТО, які можуть бути використані для попередження несправностей на ранніх етапах. Це дає змогу знизити ймовірність серйозних поломок та забезпечити безперервність роботи обладнання, що в свою чергу веде до зменшення витрат на ремонт та ТО.

Завдяки високій точності моделі дерево рішень може використовуватися для оцінки ефективності різних технічних заходів, спрямованих на запобігання поломок. Наприклад, модель може допомогти визначити, які з прийнятих рішень дійсно сприяють зниженню частоти поломок, а які не мають значного впливу. Це дозволяє на основі конкретних даних приймати більш обґрунтовані рішення щодо подальших інвестицій у технічне обслуговування або модернізацію обладнання.

Проте варто зазначити, що, як і будь-яка інша модель, дерево рішень не є бездоганним: у разі неповних або помилкових даних для навчання вона може робити неточні прогнози. Важливо також враховувати, що модель потребує якісних й репрезентативних даних для навчання, оскільки від цього безпосередньо залежить її здатність до точного прогнозування. Незважаючи на ці обмеження, дерево рішень є надзвичайно потужним інструментом, який можна використовувати для підвищення надійності обладнання і зниження ймовірності поломок.

Одним із важливих аспектів цієї моделі є її висока точність у прогнозуванні ймовірностей поломок, яка дозволяє надійно передбачити, чи є обладнання схильним до несправностей, що дає змогу операторам вчасно вжити необхідних заходів і уникнути серйозних поломок. Це, в свою чергу, дозволяє значно знизити витрати на ремонти та забезпечити більш ефективне використання ресурсів.

Окрім точності, важливим показником ефективності моделі є

ROC-крива, вона відображає здатність моделі правильно класифікувати об'єкти на позитивні та негативні випадки в залежності від обраного порогу. Модель з високою ROC-кривою демонструє здатність правильно передбачати поломки на всіх рівнях порогу, що свідчить про її високу ефективність. Відповідно, дерево рішень, яке має високу ROC-криву, може служити надійним інструментом для прогнозування ймовірностей поломок та забезпечення стабільності в роботі обладнання.

Модель випадкових лісів на основі змодельованих даних (рисунок 4.5) демонструє високу ефективність у прогнозуванні ймовірностей поломок. Точність моделі становить 0,984, це означає, що вона правильно класифікує 98,4% випадків. Цей показник є дуже високим й свідчить про те, що модель здатна майже безпомилково визначати ймовірність поломок, а це є критично важливим для забезпечення надійності роботи обладнання.

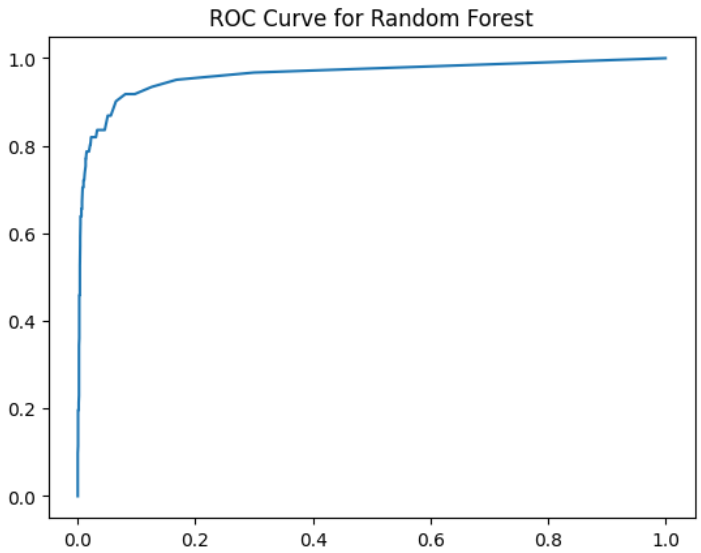


Рисунок 4.5 – ROC-крива для випадкового лісу

Такий високий рівень точності говорить про те, що модель випадкових лісів добре справляється з великими обсягами даних і здатна виявляти складні патерни, які можуть бути важко помітними для інших моделей. Це дозволяє використовувати її для раннього виявлення поломок, що, у свою

чергу, допомагає планувати технічне обслуговування та мінімізувати витрати на ремонт.

Випадкові ліси мають також перевагу у вигляді здатності обробляти високовимірні дані та добре справляються з шумами та викидами, що можуть виникнути при зборі і аналізі даних. Це робить модель випадкових лісів надійним інструментом для прогнозування ймовірностей поломок у складних системах, де можуть бути наявні різноманітні варіації в даних.

ROC-крива моделі випадкових лісів також має дуже високі значення, що свідчить про високу здатність моделі точно прогнозувати ймовірності поломок. Цей показник є критично важливим для оцінки ефективності моделі, оскільки він демонструє, як добре модель здатна відрізняти між різними класами (наприклад, поломка чи не поломка) при різних порогових значеннях.

З високим значенням ROC-кривої можна зробити висновок, що модель випадкових лісів є надзвичайно надійним інструментом для прогнозування ймовірностей поломок. Її висока точність та здатність до виявлення ймовірностей на ранніх етапах дозволяють виявляти проблеми в їхніх початкових стадіях. Це дає змогу оперативно реагувати на потенційні поломки, що, в свою чергу, може запобігти виникненню серйозних і дорогих поломок або збоїв у роботі обладнання. Таким чином, модель може бути використана для планування своєчасного ТО, що покращує ефективність та надійність експлуатації системи.

Модель випадкових лісів є інструментом, який може використовуватися для виявлення поломок на основі різноманітних даних, що збираються з датчиків, даних експлуатації та ТО. Ця модель здатна прогнозувати ймовірності поломок, що дозволяє своєчасно виявляти проблеми і вживати необхідних заходів для їхнього усунення. Це може значно підвищити ефективність планування ТО, адже за допомогою таких прогнозів можна мінімізувати ризики серйозних поломок і відповідних витрат на ремонт.

Використання моделі також дає можливість оцінити ефективність заходів, спрямованих на запобігання поломок. Наприклад, можна перевірити, чи допомагають конкретні заходи по зниженню зносу інструментів або покращенню умов експлуатації зменшити ймовірність поломок.

Проте варто зауважити, що модель випадкових лісів, хоча й є потужним інструментом, не є бездоганною. Вона може робити помилки, зокрема, у випадках, коли набір даних для навчання є неповним або неякісним. Тому важливо забезпечити якісні та актуальні дані для навчання моделі.

Особливо важливим аспектом є точність моделі в прогнозуванні ймовірностей поломок. Це дозволяє з високою ймовірністю визначати, чи є обладнання схильним до поломки, що може допомогти операторам уникнути серйозних поломок і, відповідно, знизити витрати на ремонтні роботи.

ROC-крива є ще одним важливим індикатором ефективності моделі. Вона дає змогу оцінити, як модель прогнозує ймовірності поломок при різних порогових значеннях, і якщо модель має високу ROC-криву, це свідчить про її здатність робити точні прогнози для всіх значень порогу.

Отже, модель випадкових лісів є ефективним і надійним інструментом для прогнозування ймовірностей поломок. Вона дозволяє підвищити надійність обладнання, запобігти серйозним технічним проблемам та оптимізувати витрати на технічне обслуговування.

Метод опорних векторів (SVM) продемонстрував високу ефективність при прогнозуванні ймовірностей поломок (рисунок 4.6). Точність моделі, яка складає 0,970, свідчить про те, що вона правильно класифікує понад 97% випадків. Це означає, що модель має хорошу здатність до розпізнавання патернів у даних й може з високою ймовірністю визначати, коли обладнання потребує уваги через можливу поломку.

ROC-крива моделі, яка також демонструє високий рівень, підтверджує, що модель ефективно прогнозує ймовірності поломок для різних значень порогових значень. Це дозволяє використовувати модель на практиці для

точного визначення, коли треба вживати заходів для запобігання поломкам.

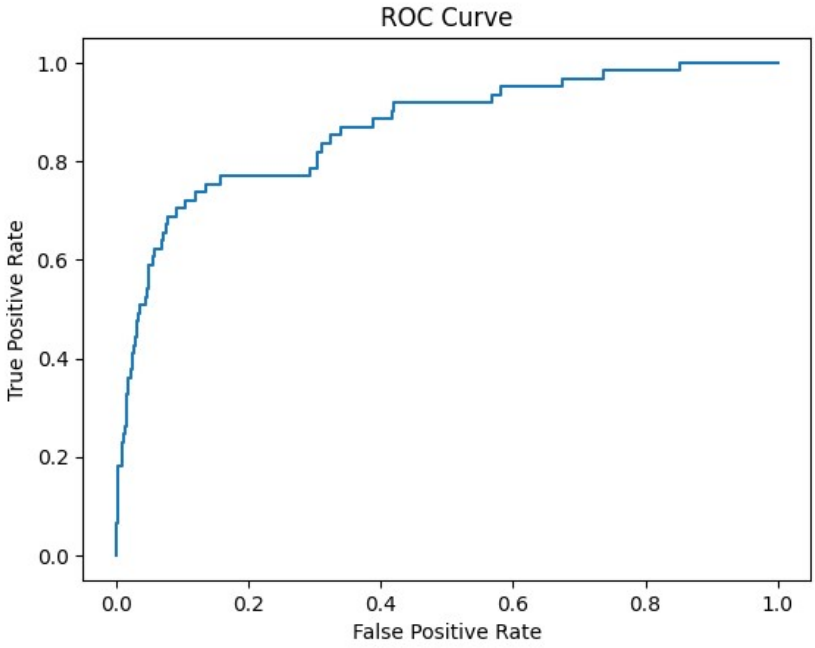


Рисунок 4.6 - ROC-крива методу опорних векторів

Модель методу опорних векторів є потужним інструментом для прогнозування ймовірностей поломок, і її можна використовувати для підвищення надійності обладнання та оптимізації планування ТО. Вона також здатна знижувати ризики серйозних поломок, що важливо для забезпечення безперервної та ефективної роботи обладнання.

На підставі цих даних можна зробити висновок, що модель методу опорних векторів є ще одним надійним інструментом для підвищення надійності обладнання та запобігання серйозним проблемам. Висока точність класифікації та позитивні результати ROC кривої свідчать про здатність моделі ефективно прогнозувати ймовірність поломок на ранніх етапах. Це дозволяє своєчасно виявляти потенційні збої та вживати заходів для їх запобігання, що значно знижує ймовірність серйозних неполадок і підвищує загальну ефективність і надійність виробничого процесу.

Модель методу опорних векторів є ефективним інструментом для

прогнозування ймовірностей поломок на основі різноманітних даних, зібраних з різних джерел, таких як датчики, дані експлуатації та інформація про технічне обслуговування. Завдяки своїм можливостям аналізувати ці дані, модель може допомогти своєчасно виявляти потенційні проблеми, що дозволяє уникнути більш серйозних несправностей у роботі обладнання.

Крім того, ця модель може бути корисною при плануванні ТО, забезпечуючи можливість розробки ефективних стратегій для запобігання поломок. Прогнозуючи ймовірності відмов, модель дає змогу створити плани обслуговування, які мінімізують ймовірність несправностей та сприяють забезпеченню стабільної роботи обладнання протягом довгого часу.

Модель також може використовуватися для оцінки ефективності заходів, спрямованих на запобігання поломок. Оцінюючи, як різні стратегії впливають на ймовірність відмов, вона дозволяє визначити, наскільки успішними є запроваджені заходи, що може допомогти оптимізувати процеси технічного обслуговування.

Попри високу точність прогнозів, слід зазначити, що модель не є безпомилковою. Вона може допускати помилки, особливо якщо дані для навчання неповні або містять помилки. Тому важливо ретельно перевіряти якість даних перед застосуванням моделі в реальних умовах. Незважаючи на ці обмеження, метод опорних векторів залишається потужним і ефективним інструментом для підвищення надійності обладнання, що дозволяє мінімізувати ризики відмов і забезпечувати безперервну роботу технологічних процесів.

4.3 Аналіз результатів тестування

На основі проведених наукових досліджень можна зробити кілька важливих висновків щодо ефективності різних методів прогнозування ймовірностей поломок. Логістична регресія, дерева рішень, випадкові ліси та метод опорних векторів продемонстрували високу ефективність у вирішенні

цієї задачі. Кожен із цих методів показав високу точність прогнозів, що свідчить про їх здатність правильно класифікувати ймовірність поломки з мінімальною кількістю помилок.

Вибір найбільш оптимального методу залежить від конкретних вимог та умов експлуатації в організації. Якщо головним критерієм є висока точність прогнозу, то варто звернути увагу на метод дерев рішень, випадкових лісів або опорних векторів, оскільки вони мають дуже високі показники точності. Ці методи здатні забезпечити надійне прогнозування з мінімальною кількістю помилок, що є критично важливим для підтримання стабільності технологічних процесів.

З іншого боку, якщо для організації важливими є простота, зрозумілість та легкість інтерпретації моделі, то логістична регресія може бути більш підходящим вибором. Вона є прозорою у своєму функціонуванні, що дозволяє легше зрозуміти механізм прийняття рішень і забезпечити зручність у використанні для практичного застосування.

Отже, вибір між цими методами прогнозування залежить від конкретних вимог до точності, складності та практичності моделі в умовах експлуатації, що дозволяє вибудовувати оптимальні стратегії ТО на основі прогнозів ймовірностей поломок.

Для досягнення максимальної ефективності в прогнозуванні ймовірностей поломок за допомогою різних моделей варто врахувати кілька ключових рекомендацій. Перша з них полягає в необхідності ретельного тестування будь-якої моделі на спеціально підготовлених тестових наборах даних. Це дозволяє оцінити її точність і надійність у прогнозуванні, що є критично важливим для визначення рівня впевненості в її прогностичних можливостях.

Наступною важливою рекомендацією є використання моделей для виявлення поломок, спираючись на різні джерела даних. Такими джерелами можуть бути показники датчиків, експлуатаційні дані, а також дані ТО. Це дозволяє отримати більш повну картину стану обладнання, що підвищує

ефективність прогнозування.

Моделі також можуть бути корисними в процесі розробки планів ТО. Їх застосування дає можливість більш точно передбачити можливі проблеми і своєчасно вжити заходів, що дозволяє уникнути поломок і таким чином значно покращити загальну надійність обладнання.

Окрім того, ці моделі можуть бути використані для оцінки ефективності вже вжитих заходів, спрямованих на запобігання поломок. Вони допомагають зрозуміти, наскільки ефективно працюють ці заходи та чи потрібно вносити корективи для досягнення кращих результатів. Використання моделей у цих цілях дозволяє здійснювати не лише прогнозування, а й постійний моніторинг та вдосконалення процесів ТО.

Слід врахувати, що, незважаючи на високу ефективність моделей для прогнозування ймовірностей поломок, вони не можуть бути визнані абсолютно універсальними рішеннями. Ці моделі здатні значно допомогти операторам у виявленні потенційних проблем і зниженні ризиків серйозних поломок, однак жодна модель не може гарантувати повну відсутність неполадок, оскільки завжди існують фактори, які не піддаються точному прогнозуванню.

Важливим аспектом є не лише точність моделей, а й інші критерії, такі як складність моделі, час, необхідний для її навчання, та доступність якісних даних для тренування. Вибір оптимальної моделі для конкретної ситуації має враховувати баланс між її точністю та іншими вимогами, зокрема, здатністю працювати в реальному часі та з різними типами даних.

Моделі прогнозування ймовірностей поломок можуть стати незамінним інструментом для підвищення ефективності ТО. Вони дозволяють виявляти можливі неполадки на ранніх етапах, що допомагає запобігти їхньому розвитку в серйозні проблеми та суттєво знизити витрати, пов’язані з ремонтом та простоєм обладнання. Однак, як показує практика, вибір оптимальної моделі має базуватися на специфічних потребах та вимогах конкретної організації. Різні виробничі умови, типи обладнання та

організаційні процеси можуть потребувати різних підходів до вибору моделі, що забезпечить найкращий результат для конкретної ситуації.

Проведене дослідження методів програмного моніторингу ТО на елеваторному комплексі дозволяє зробити кілька важливих висновків щодо ефективності застосування різних технологій для прогнозування ймовірностей поломок та підвищення надійності обладнання. Використання сучасних методів машинного навчання, таких як логістична регресія, дерева рішень, випадкові ліси та методи опорних векторів, продемонструвало високу точність у прогнозуванні поломок, що дає змогу знижувати ризики виникнення серйозних проблем і значно оптимізувати процеси ТО.

Завдяки застосуванню програмного моніторингу можна своєчасно виявляти можливі поломки на ранніх етапах, що дозволяє не лише забезпечити безперервну роботу комплексу, а й суттєво зменшити витрати на аварійні ремонти та зупинки. Моделі прогнозування поломок, незважаючи на свою високу точність, не є ідеальними, тому їх використання повинно бути доповнене іншими факторами, такими як доступність даних, складність моделей та час, необхідний для їх тренування.

Завдяки цим інструментам, можна не тільки покращити процеси ТО, а й розробити ефективні стратегії запобігання поломок, що дозволяє підвищити загальну ефективність і надійність роботи елеваторного комплексу. Таким чином, застосування методів програмного моніторингу і прогнозування в поєднанні з традиційними методами ТО може значно поліпшити управління та експлуатацію обладнання, що в свою чергу сприяє збільшенню продуктивності та зниженню витрат.

ВИСНОВКИ

У кваліфікаційній роботі виконано комплексне дослідження, спрямоване на підвищення ефективності обслуговування технологічного обладнання завдяки використанню сучасних підходів до аналізу даних і прогнозування. У ході роботи розглянуто існуючі методи аналізу даних, проведено їх класифікацію та оцінено з точки зору відповідності критеріям ефективності для задач ТО. Особливий акцент зроблено на таких підходах, як логістична регресія, дерева рішень, випадкові ліси та метод опорних векторів.

Здійснено ретельний аналіз критеріїв якості прогнозуючих моделей, що включає точність, повноту, AUC-ROC показники та стабільність роботи моделей. Це дало змогу зробити обґрунтований вибір методів машинного навчання для розробки системи моніторингу технічного обслуговування, зважаючи на їхню точність, адаптивність та здатність працювати з великими наборами даних.

Розроблена архітектура системи збору та аналізу даних є багаторівневою структурою, що забезпечує інтеграцію різних джерел даних, включаючи датчики, інформацію про експлуатацію обладнання та технічне обслуговування. Використання сучасних технологій, таких як бази даних і платформи для машинного навчання, дозволило створити систему, здатну забезпечувати постійний моніторинг стану обладнання та надавати точні прогнози щодо його поломок.

Реалізація методів програмного моніторингу передбачала повний цикл обробки даних, від збору до побудови моделей машинного навчання. У процесі роботи розроблено ефективні алгоритми для прогнозування ймовірностей поломок, засновані на глибоких нейронних мережах, таких як LSTM і GRU. Ці моделі враховують часову динаміку даних, що робить їх особливо корисними для аналізу експлуатаційних показників обладнання на

елеваторному комплексі.

Експериментальні дослідження, проведені на основі реальних даних елеваторного комплексу, продемонстрували високу точність запропонованих моделей, зокрема, точність прогнозів для логістичної регресії склала 97.4%, для дерев рішень - 97.7%, для випадкового лісу - 98.4%, а для методу опорних векторів - 97%. Крім того, оцінка ефективності методів показала їх здатність виявляти потенційні поломки на ранніх стадіях, що є ключовим фактором для забезпечення надійності роботи обладнання.

Результати аналізу підтвердили, що впровадження системи програмного моніторингу дозволяє суттєво зменшити ймовірність виникнення аварійних ситуацій, оптимізувати графік ТО та скоротити витрати на ремонт. Це забезпечує значне підвищення економічної ефективності роботи елеваторного комплексу та знижує ризики, пов’язані з раптовими відмовами обладнання.

В роботі запропоновано наступні методи програмного моніторингу ТО: збору та обробки даних телеметрії, прогнозування залишкового ресурсу, виявлення аномалій у роботі обладнання та класифікації стану обладнання.

Таким чином, виконане дослідження демонструє перспективність використання розроблених методів машинного навчання для вирішення задач моніторингу ТО на елеваторних комплексах. Отримані результати можуть стати основою для подальшого вдосконалення систем прогнозування, що дозволить не лише підвищити надійність обладнання, а й забезпечити більш ефективне використання ресурсів.

За темою роботи опубліковано тези доповіді в рамках всеукраїнської науково-практичної конференції здобувачів вищої освіти та молодих учених «Комп’ютерно-інтегровані технології автоматизації технологічних процесів на транспорті та у виробництві» [19] та на дванадцятій міжнародній науково-технічній конференції “Проблеми інформатизації” [20] (додаток Б).

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Елеваторна промисловість: традиції та інновації. Вітчизняний та світовий досвід [Електронний ресурс] : наук.-допом. бібліогр. покажч. / [упоряд. Т. П. Фесун] ; Нац. ун-т харч. технол., Наук.-техн. б-ка. – Київ, 2021. – 180 с.

2. Колотило Д. М. Технологічні процеси галузей промисловості: Навч.посібник/ А.Т. Соколовський, С.В. Гарбуз — К.: КНЕУ, 2003. — 380 с.

3. Нагорний А.В. Автоматизація технологічних процесів і систем автоматичного керування / А.В. Нагорний, В.М. Манжара: НМЦ, 2003. – 82 с.

4. Орлов О. Планування діяльності промислового підприємства: Підручник для студ. ВНЗ. — К. : Видавничий дім "Скарби", 2002. — 336 с.

5. Data Analysis Methods and Techniques [Електронний ресурс]. – Режим доступу: https://www.datapine.com/blog/data-analysis-methods-and-techniques/ – 10.11.2024 г. – Загл. з екрану.

6. Logistic Regression [Електронний ресурс]. – Режим доступу: https://www.geeksforgeeks.org/understanding-logistic-regression/ – 12.11.2024 г. – Загл. з екрану.

7. Decision Tree [Електронний ресурс]. – Режим доступу: https://www.geeksforgeeks.org/decision-tree/ – 14.11.2024 г. – Загл. з екрану.

8. Random forest [Електронний ресурс]. – Режим доступу: https://www.wikiwand.com/uk/Random_forest – 15.11.2024 г. – Загл. з екрану.

9. Support Vector Machine Algorithm [Електронний ресурс]. – Режим доступу: https://www.geeksforgeeks.org/support-vector-machine-algorithm/ – 18.11.2024 г. – Загл. з екрану.

10. ROC and AUC [Електронний ресурс]. – Режим доступу: https://developers.google.com/machine-learning/crash-course/classification/roc-andauc – 18.11.2024 г. – Загл. з екрану.

11. AUC ROC Curve [Електронний ресурс]. – Режим доступу:

https://www.geeksforgeeks.org/auc-roc-curve/ – 19.11.2024 г. – Загл. з екрану.

12. Machine (Predictive Maintenance) Classification Dataset [Електронний ресурс]. – Режим доступу: https://www.kaggle.com/datasets/shivamb/mm-classification/data – 21.11.2024 г. – Загл. з екрану.

13. Google Colaboratory —хмарне середовище розробки. [Електронний ресурс]. – Режим доступу: https://colab.research.google.com – 23.11.2024 г. – Загл. з екрану.

14. Підручник з MySQL [Електронний ресурс]. – Режим доступу: https://www.w3schools.com/MySQL/ – 25.11.2024 г. – Загл. з екрану.

15. SCADA Citect [Електронний ресурс]. – Режим доступу: https://www.aveva.com/en/products/plant-scada/ – 27.11.2024 г. – Загл. з екрану.

16. Програмовані контролери для систем керування. Навчальний посібник [Текст] / Г.І. Загарій, Н.О. Ковзель, В.С. Коновалов та ін. - Х.: ХФІ "Транспорт України", 2022. - 264 с.

17. Що таке протокол Modbus і як він працює? [Електронний ресурс]. – Режим доступу: https://dusuniot.com/uk/blog/what-is-the-modbus-protocol-and-how-does-it-work/ – 01.12.2024 г. – Загл. з екрану.

18. Price M. J. C# 9 and .NET 5 - Modern Cross-Platform Development / Mark Price. – Birmingham, 2020. – 882 с. – (Packt Publishing). – ISBN 9781800568105.

19. Мостовий А.В., наук. керівник Піскарьов О.М. Методи програмного моніторингу технічного обслуговування на елеваторному комплексі / Комп’ютерно-інтегровані технології автоматизації технологічних процесів на транспорті та у виробництві. Матеріали всеукраїнської науково-практичної конференції. – Харків, ХНАДУ, 2024. – С. 213-216.

20. Мостовий А.В. наук. керівник Піскарьов О.М. Актуальність вдосконалення методів програмного моніторингу технічного обслуговування на елеваторному комплексі / Проблеми інформатизації. Тези доповідей XII міжнародної науково-технічної конференції 21 – 22 листопада 2024 р. Том 2 (Секція 4) – Харків: ХНУРЕ, 2024 – С.62.