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)

Модель мережної аналітики з використанням технологій машинного навчання

(тема)

Виконав:

студент II курсу, групи КСМм-22-2

Лушпа Б.Є.
(прізвище, ініціали)

Спеціальність

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

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

Освітня програма

Комп’ютерні системи та мережі
(повна назва освітньої програми)

Керівник: доц. Голубничий Д.Ю.
(посада, прізвище, ініціали)

Допускається до захисту

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

2024 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 Комп'ютерні системи та мережі
(повна назва)

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

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенту Лушпі Богдану Євгеновичу
(прізвище, ім'я, по батькові)

1. Тема роботи Модель мережної аналітики з використанням технологій машинного навчання

затверджена наказом по університету від " 06 " листопада 2023 р. № 1298 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 15 січня 2024 р.

3. Вхідні дані до роботи 1) об'єкт дослідження – стільникова мережа 5G;
2) застосування методів машинного навчання та глибокого навчання;
3) інтеграція алгоритмів ML/DL до застосунків NTMA.

4. Перелік питань, що потрібно опрацювати у роботі
1) огляд технологій машинного навчання та їхнього застосування;
2) аналіз особливостей задачі NTMA;
3) аналіз особливостей стільникових мереж 5G;
4) розробка моделі;
5) проведення експериментальних досліджень;
6) висновки.

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

Слайд-презентація – 12 слайдів

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 технологій машинного навчання та	07.11.23-13.11.23	
	їхнього застосування		
2	Аналіз особливостей задачі NTMA	14.11.23-20.11.23	
3	Аналіз особливостей стільникових мереж 5G	21.11.23-23.11.23	
4	Розробка моделі	24.11.23-06.12.23	
5	Проведення експериментів	07.12.23-23.12.23	
6	Оформлення матеріалів кваліфікаційної роботи	26.12.23-02.01.24	
7	Подання кваліфікаційної роботи керівникові та	03.01.24-06.01.24	
	її попередній захист		
8	Подання кваліфікаційної роботи на рецензу-	09.01.24-12.01.24	
	вання		

Дата видачі завдання 06 листопада 2023 р.

Студент _____
(підпис)

Керівник роботи _____ доц. Голубничий Д.Ю.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 78 с., 19 рис., 5 табл., 2 дод., 27 джерел.

КЕРУВАННЯ МЕРЕЖЕЮ, МАШИННЕ НАВЧАННЯ, МЕРЕЖІ 5G, МОНІТОРИНГ І АНАЛІЗ МЕРЕЖНОГО ТРАФІКУ, ХЕНДОВЕР, NTMA, NWDAF, QOS.

Метою кваліфікаційної роботи є розробка моделі системи мережної аналітики, яка використовує методи машинного навчання.

Основні внески роботи включають таке: розглянуто традиційні методики, засновані на навчанні, для NTMA; проаналізовано ключові характеристики та застосування NTMA; розглянуто методи ML, які використовуються у застосунках NTMA; розглянуто NWDAF в архітектурі стільникових мереж 5G; запропоновано модель системи для інтелектуальної мережної аналітики у стільникових мережах 5G. Використано кілька методів ML для вирішення двох основних проблем: прогнозування навантаження на мережу за допомогою аналізу часових рядів, зокрема за допомогою моделей лінійної регресії, LSTM і RNN; класифікація аномалій в мережі за допомогою моделей логістичної регресії і метода підсилювання градієнта XGBoost.

Результаті цієї роботи призначені для фахівців у сфері комунікаційних систем і мереж, які планують використовувати аналітичні системи на основі штучного інтелекту для комунікаційних інфраструктур.

ABSTRACT

Master’s thesis: 78 pages, 19 figures, 5 tables, 2 appendices, 27 sources.

5G NETWORKS, HANDOVER, MACHINE LEARNING, NETWORK MANAGEMENT, NETWORK TRAFFIC MONITORING AND ANALYSIS, NTMA, NWDAF, QOS.

The major goal of this thesis is to develop a network analytics system model that uses machine learning methods.

The main contributions of the paper include the following: traditional learning-based techniques for NTMA are reviewed; the key characteristics and applications of NTMA are analyzed; ML methods used in NTMA applications are reviewed; NWDAF in the architecture of 5G cellular networks are considered; a system model for intelligent network analytics in 5G cellular networks is proposed. Several ML methods have been used to solve two main problems: forecasting network load using time series analysis, particularly using linear regression, LSTM, and RNN models; classification of anomalies in the network using logistic regression models and the XGBoost gradient boosting method.

The results of this work are intended for experts in the field of communication systems and networks who plan to use analytical systems based on artificial intelligence for communication infrastructures.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

3GPP – Проект партнерства третього покоління (англ., 3rd Generation Partnership Project)

AF – прикладна функція (англ., Application Function)

AI – штучний інтелект (англ., Artificial Intelligence)

AMF – управління доступом і мобільністю (англ., Access and Mobility Management)

ANN – штучна нейронна мережа (англ., Artificial Neural Network)

AUC – площа під кривою (англ., Area Under Curve)

AUC-ROC – область під робочими характеристиками приймача (англ., Area Under Receiver Operating Characteristics)

DL – глибоке навчання (англ., Deep Learning)

DPI – глибока перевірка пакетів (англ., Deep Packet Inspection)

GAN – генеративні змагальні мережі (англ., Generative Adversarial Networks)

H2H – «людина-людина» (англ., Human-to-Human)

IoT – Інтернет речей (англ., Internet of Things)

LogReg – логістична регресія (англ., Logistic Regression)

LR – лінійна регресія (англ., Linear Regression)

LSTM – довга короткочасна пам’ять (англ., Long Short-Term Memory)

M2M – «машина-машина» (англ., Machine-to-Machine)

MAE – середня абсолютна похибка (англ., Mean Absolute Error)

MAPE – середня абсолютна похибка у відсотках (англ., Mean Absolute Percentage Error)

ML – машинне навчання (англ., Machine Learning)

NF – мережна функція (англ., Network Function)

NTMA – моніторинг та аналіз мережного трафіка (англ., Network

Traffic Monitoring and Analysis)

NWDAF – функція аналізу мережних даних (англ., Network Data Analytics Function)

QoE – якість досвіду (англ., Quality of Experience)

QoS – якість обслуговування (англ., Quality of Service)

RAN – мережа радіодоступу (англ., Radio Access Network)

RNN – рекурсивна нейронна мережа (англ., Recursive Neural Network)

RRC – керування радіоресурсами (англ., Radio Resource Control)

SA – автономний (англ., Standalone)

SBA – сервісна архітектура (англ., Service-Based Architecture)

SBI – сервісний інтерфейс (англ., Service-Based Interface)

SLA – рівень обслуговування (англ., Service Level Agreement)

SPI – поверхнева перевірка пакетів (англ., Shallow Packet Inspection)

SubsCat – категорія абонента (англ., Subscriber Category)

TCP/IP – протокол керування передачею та Інтернет-протокол (англ., Transmission Control Protocol – Internet Protocol)

UE – обладнання користувача (англ., User Equipment)

XGBoost – екстремальне підсилювання градієнта (англ., eXtreme Gradient Boosting)

ВСТУП

У сучасному технологічному середовищі, яке розвивається дуже швидко, мережні операції відіграють вирішальну роль у забезпеченні надійного підключення та високої продуктивності. Однак через експоненційне зростання обсягів даних і складності мереж традиційних ручних підходів до мережних операцій стає недостатньо. З цього виникає потреба у машинному навчанні (англ., ML – Machine Learning) і аналізі даних. Отже, актуальною є проблема застосування ML і аналітики даних у мережних операціях з підкресленням їхнього потенціалу кардинально змінити спосіб керування та оптимізації мереж.

Сучасні комунікаційні системи та мережі, наприклад, Інтернет речей (англ., IoT – Internet of Things) і стільникові мережі, генерують величезну та різнорідну кількість даних трафіку. У таких мережах традиційні методи моніторингу та аналізу даних стикаються з деякими проблемами, наприклад, з точністю та ефективною обробкою великих даних у режимі реального часу. Крім того, модель мережного трафіку, насамперед в стільникових мережах, показує дуже складну поведінку через різні фактори, такі як мобільність пристроїв і неоднорідність мережі. Для полегшення аналітики та виявлення знань у системах великих даних для розпізнавання прихованих і складних закономірностей ефективно використовується глибоке навчання (англ., DL – Deep Learning). Дослідники в області мереж застосовують моделі глибокого навчання для програм моніторингу та аналізу мережного трафіку (англ., NTMA – Network Traffic Monitoring and Analysis), наприклад, з метою класифікації та прогнозування трафіку.

Стільникові мережі 5G мають багато нових функцій порівняно із застарілими мережами стільникового зв’язку, наприклад функцію аналізу мережних даних (англ., NWDAF – Network Data Analytics Function), яка дозволяє адміністраторам мереж або впроваджувати власні методології

аналізу даних на основі машинного навчання, або інтегрувати до своїх мереж рішення сторонніх розробників.

В роботі спочатку проводиться огляд протоколів NWDAF, визначених в стандартних документах 3GPP. Далі генерується синтетичний набір даних для клітинок мереж 5G на основі полів, визначених специфікаціями 3GPP. При цьому, до набору даних додаються деякі аномалії (наприклад, раптове збільшення трафіку в конкретній комірці), після чого ці аномалії класифікуються по кожній комірці, категорії абонентів і обладнанню користувача. Далі, для вивчення оцінки інформації про поведінку (наприклад, аномалії в мережному трафіку) і можливості прогнозування мережного навантаження NWDAF, реалізуються три моделі ML, а саме лінійна регресія, довгострокова пам’ять і рекурсивні нейронні мережі.

Для прогнозування навантаження на мережу використовуються три різні моделі з метою мінімізації середньої абсолютної похибки, яка обчислюється шляхом віднімання фактично згенерованих даних із прогнозованого значення моделі. Для класифікації аномалій використовуються дві моделі ML для збільшення площі під кривою робочих характеристик приймача, а саме логістична регресія та екстремальне градієнтне підсилювання. Згідно з результатами моделювання, алгоритми нейронної мережі перевершують лінійну регресію у прогнозуванні навантаження на мережу, тоді як алгоритм підсилювання градієнта на основі дерева перевершує логістичну регресію у виявленні аномалій. Очікується, що ці оцінки підвищать продуктивність мережі 5G через NWDAF.

1 МАШИННЕ НАВЧАННЯ ТА АНАЛІЗ ДАНИХ ДЛЯ МЕРЕЖНИХ ОПЕРАЦІЙ

1.1 Загальні відомості

Машинне навчання та аналітика даних у мережних операціях включають використання прогресивних обчислень і фактичних моделей для вилучення інформації з мережної інформації. Алгоритми ML (рисунок 1.1) вивчають історичні дані для того, щоб робити прогнози, виявляти закономірності та автоматизувати процеси прийняття рішень. Аналітика даних зосереджена на аналізі та інтерпретації мережних даних для отримання цінної інформації та підтримки прийняття обґрунтованих рішень.

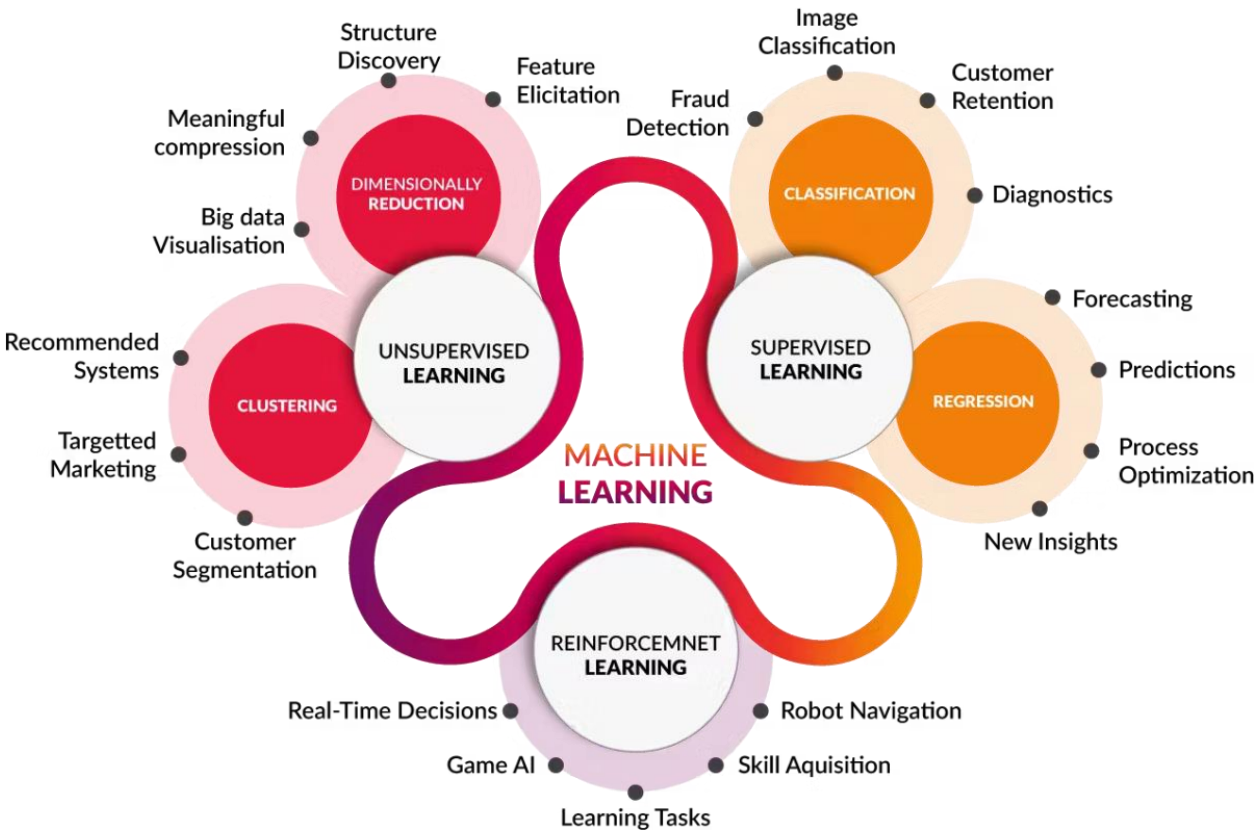


Рисунок 1.1 – Машинне навчання

Ці підходи відрізняються від традиційних методів керування мережею, пропонуючи автоматизовані та інтелектуальні рішення, які можуть впоратися зі складністю та масштабом сучасних мереж. ML і аналітика даних надають змогу мережним адміністраторам проактивно контролювати та вдосконалювати свої системи, що забезпечує прогресивне виконання, підвищену надійність і розширену безпеку.

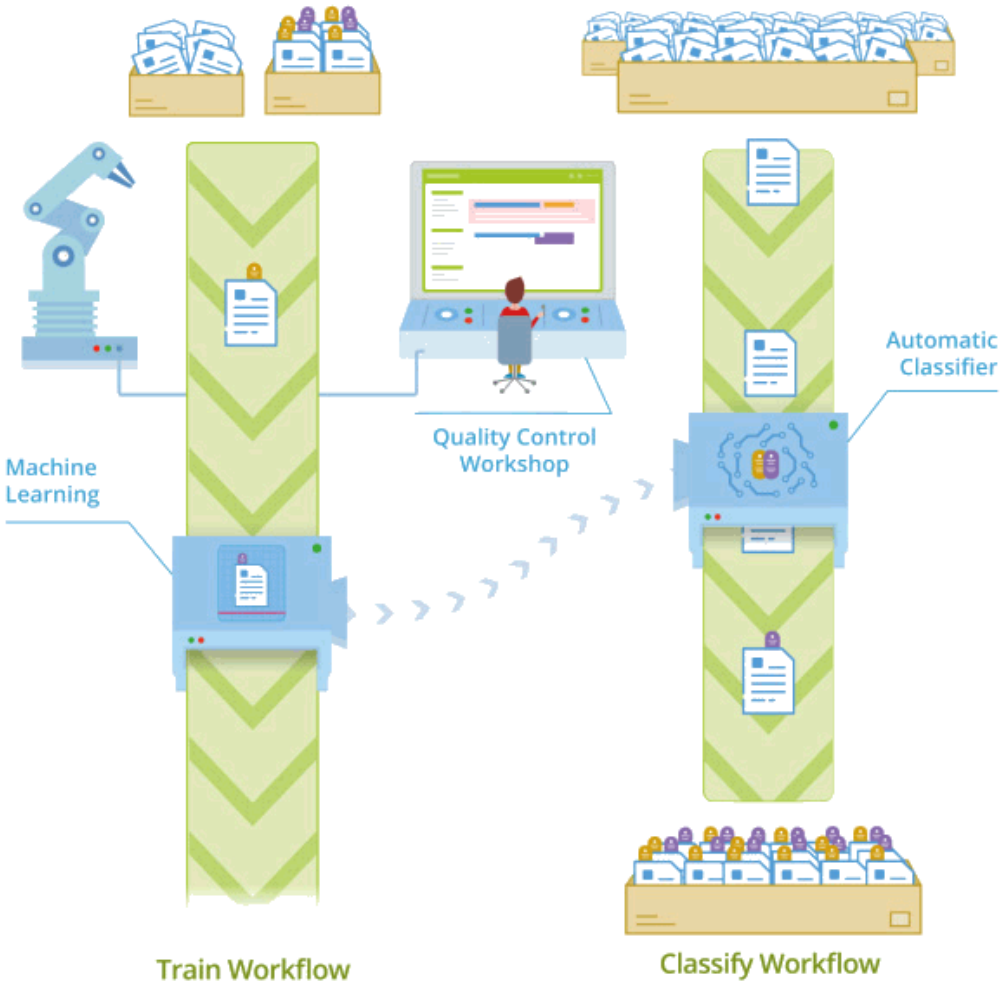


Рисунок 1.2 – Робочий процес машинного навчання

Адміністратори мереж стикаються зі значними проблемами в управлінні великомасштабними мережами. Традиційні «ручні» підходи займають багато часу, схильні до помилок і не встигають за динамічною природою мережних ситуацій. Для вирішення цих проблем ML і аналітика

даних пропонують автоматизовані та інтелектуальні підходи. Використовуючи алгоритми ML, мережні адміністратори можуть завчасно виявляти та вирішувати проблеми з продуктивністю мережі. Прогнозна аналітика дозволяє планувати потужності та розподіляти ресурси на основі історичних даних і майбутніх прогнозів попиту. Оптимізація мережі в режимі реального часу, керована ідеями машинного навчання, дозволяє динамічно коригувати мережні конфігурації, маршрутизацію та керування трафіком [1].

1.2 Застосування машинного навчання та аналізу даних у мережних операціях

1.2.1 Оптимізація продуктивності мережі

ML і аналітика даних відіграють вирішальну роль в оптимізації продуктивності мережі. Завдяки постійному аналізу мережних даних алгоритми ML можуть виявляти потенційні вузькі місця, прогнозувати перевантаження мережі та рекомендувати оптимальні стратегії маршрутизації. Цей проактивний підхід (рисунок 1.3) дозволяє мережним адміністраторам вирішувати проблеми продуктивності до того, як вони вплинуть на кінцевих користувачів.

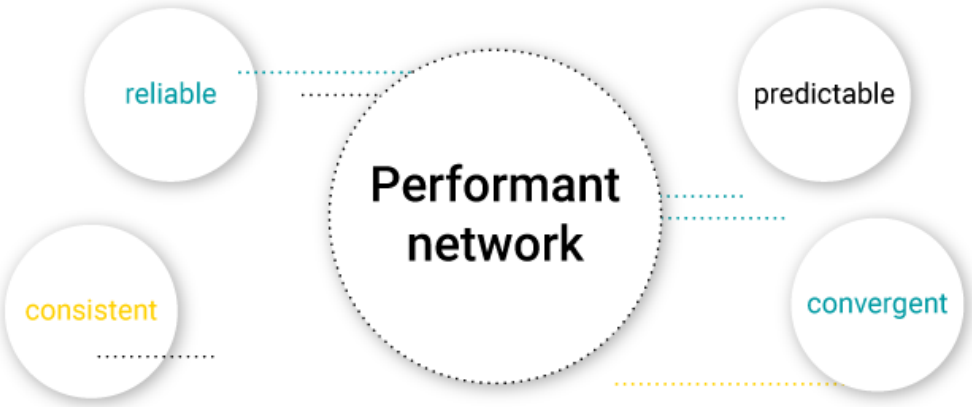


Рисунок 1.3 – Продуктивність мережі машинного навчання

Крім того, алгоритми ML можуть використовуватися для аналізу історичних даних продуктивності, що надає змогу виявляти закономірності та аномалії, забезпечуючи прогнозне технічне обслуговування та завчасне оновлення. Використовуючи ці можливості, мережні адміністратори можуть підвищити продуктивність мережі, скоротити час простою та покращити загальну роботу клієнта.

1.2.2 Виявлення та усунення несправностей

Збої в мережі можуть суттєво вплинути на її роботу. ML і аналітика даних пропонують потужні інструменти для автоматичного виявлення несправностей і їхнього усунення. Для виявлення аномалій і ненормальної поведінки мережі алгоритми ML можуть аналізувати мережні дані в режимі реального часу [1]. Це дозволяє мережним адміністраторам вчасно виявляти та ізолювати потенційні проблеми.

Крім того, алгоритми ML можуть виконувати аналіз першопричини шляхом кореляції даних з різних мережних пристроїв і систем. Це прискорює процес усунення несправностей, зменшує середній час ремонту (англ., MTTR – Mean Time to Repair) і мінімізує час простою мережі [2]. Прогнозне технічне обслуговування на основі аналізу ML також допомагає запобігти збоям мережі, виявляючи потенційні проблеми до їхнього загострення.

1.2.3 Аналіз загроз та безпека

Безпека мережі є першочерговою турботою для мережних адміністраторів. ML і аналітика даних значно сприяють посиленню безпеки мережі та можливостей аналізу загроз. Алгоритми ML можуть аналізувати моделі мережного трафіку, виявляти аномалії та потенційні загрози безпеці [3].

Використовуючи виявлення загроз на основі ML, мережні

адміністратори можуть швидко реагувати на порушення безпеки та запобігати несанкціонованому доступу до мережі. Системи виявлення вторгнень на основі ML можуть аналізувати поведінку мережі та виявляти потенційні атаки чи підозрілі дії, підвищуючи загальну безпеку мережі. Передові методи аналітики також дозволяють ідентифікувати шаблони та тенденції в безпеці мережі, допомагаючи ранньому виявленню нових загроз.

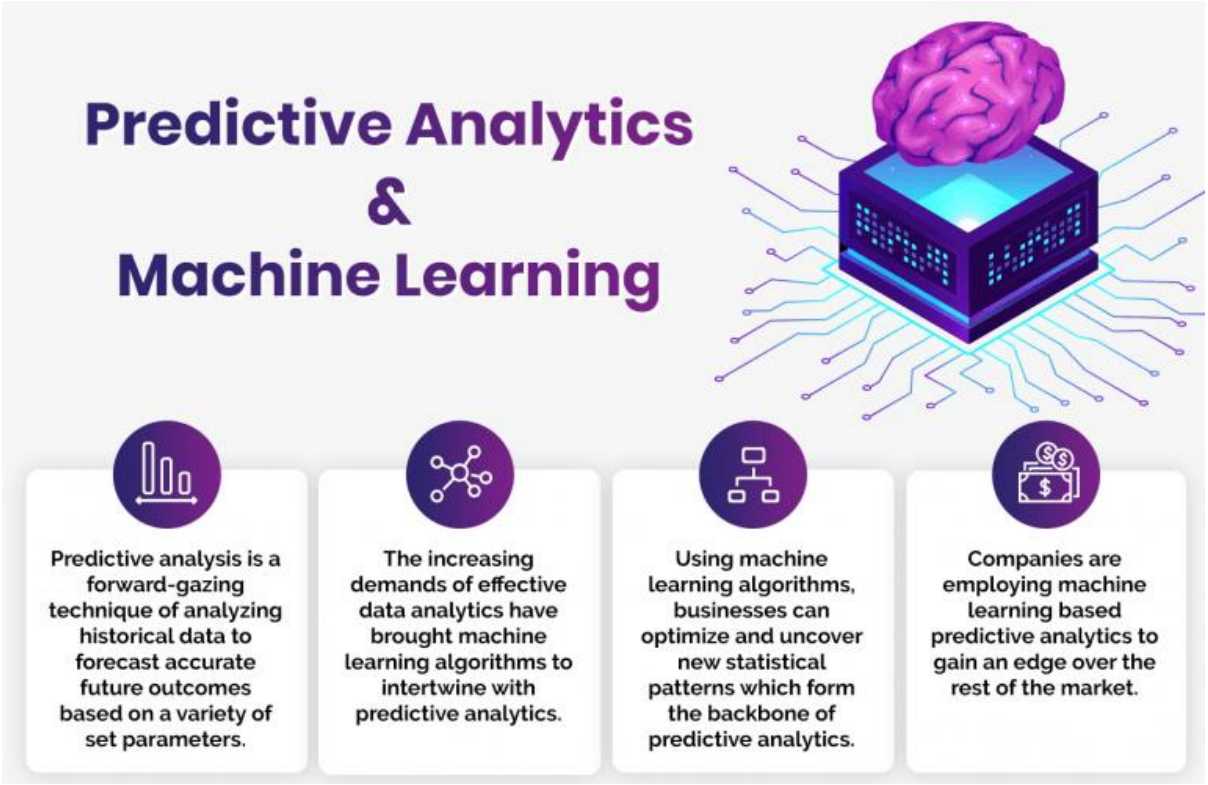


Рисунок 1.4 – Машинне навчання та прогнозний аналіз

1.3 Переваги та проблеми ML і аналізу даних у мережних операціях

1.3.1 Переваги

Застосування ML і аналізу даних у мережних операціях надає кілька переваг (рисунок 1.4).

Підвищена продуктивність і надійність мережі. Алгоритми ML оптимізують мережні конфігурації та розподіл ресурсів. Це сприяє

підвищенню продуктивності та зменшенню затримки. Прогнозна аналітика допомагає виявити потенційні проблеми до того, як вони вплинуть на мережу, забезпечуючи більшу надійність.

Підвищення операційної ефективності та економія коштів. ML і аналітика даних автоматизують рутинні завдання, зменшуючи ручні зусилля та дозволяючи адміністраторам мереж зосередитися на стратегічних ініціативах. Це спрощує мережні операції, підвищує ефективність і призводить до економії коштів за рахунок оптимізації використання ресурсів.

Швидше усунення несправностей і вирішення проблем. Алгоритми ML можуть швидко виявляти та ізолювати несправності мережі, прискорюючи процес усунення несправностей. Це скорочує MTTR і мінімізує час простою мережі, що сприяє підвищенню задоволеності клієнтів.

Покращена безпека та пом’якшення загроз. Рішення безпеки на основі ML забезпечують виявлення та запобігання загрозам у реальному часі, посилюючи безпеку мережі. Завчасно виявляючи потенційні порушення безпеки, мережні адміністратори можуть негайно вжити заходів для зменшення ризиків і захисту конфіденційних даних.

1.3.2 Проблеми

Зважаючи на те, що ML і аналітика даних пропонують значні переваги, їхнє впровадження в мережних операціях також створює певні проблеми.

Якість, цілісність і доступність даних. Алгоритми ML вимагають високоякісних і надійних даних для точних прогнозів і розуміння. Забезпечення цілісності, узгодженості та доступності даних у різноманітних джерелах мережі може бути складним завданням, що потребує надійних процесів керування даними та інфраструктури керування даними.

Конфіденційність і етичні міркування. Мережні дані часто містять конфіденційну інформацію, що викликає занепокоєння щодо конфіденційності. Адміністратори мереж повинні обробляти та обробляти

дані відповідно до правил конфіденційності та етичних принципів, щоб захистити конфіденційність користувачів і зберегти довіру.

Інтеграція ML і аналітики в існуючу мережну інфраструктуру. Інтеграція можливостей ML і аналітики даних в існуючу мережну інфраструктуру може бути складною. Адміністраторам мереж необхідно враховувати такі фактори, як сумісність, масштабованість та інтеграція даних, щоб безперешкодно впроваджувати рішення на основі ML у свою роботу.

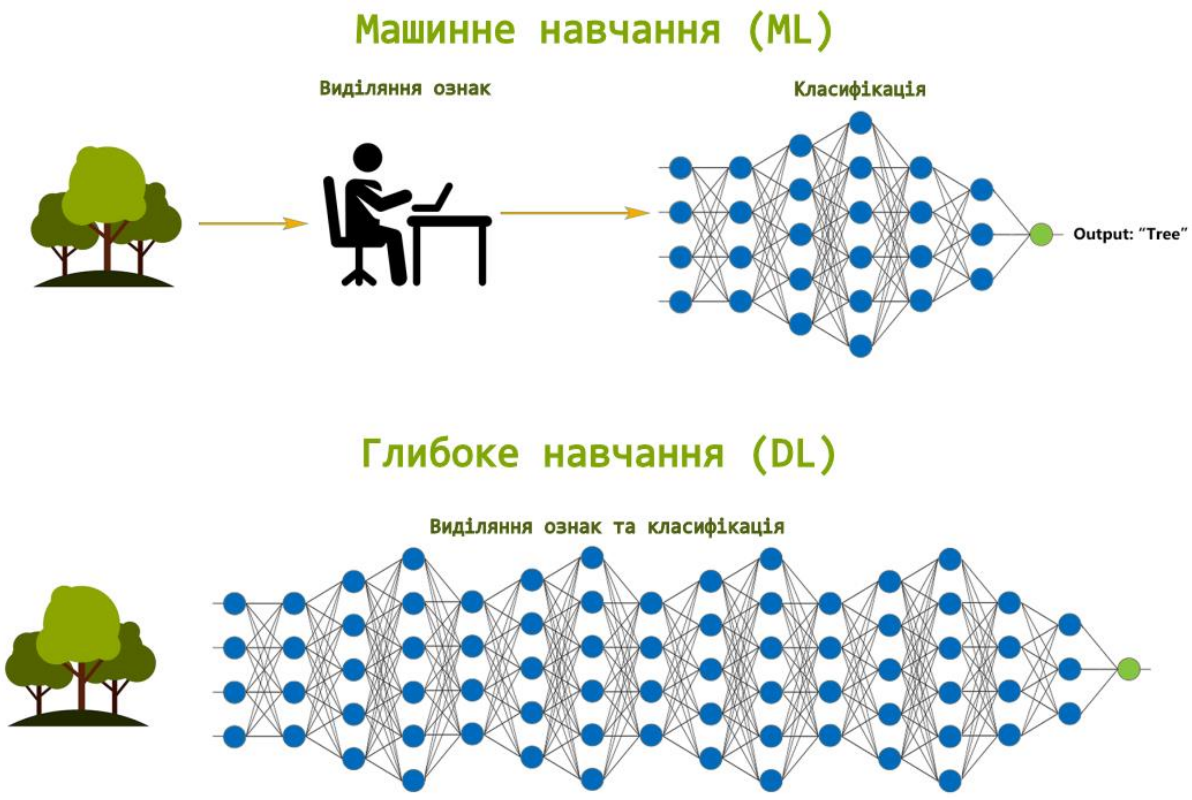


Рисунок 1.5 – Порівняння ML і DL

Навички та досвід, необхідні для ефективного впровадження. ML та аналітика даних вимагають спеціальних навичок і досвіду. Адміністраторам мереж потрібні професіонали, які розуміють як принципи мереж, так і алгоритми машинного навчання, щоб ефективно впроваджувати та керувати мережними операціями на основі машинного навчання.

1.4 Існуючі дослідження та практичні приклади

1.4.1 Оптимізація продуктивності мережі на основі ML

Під час масштабного розгортання мережі телекомунікаційна компанія використала алгоритми ML для оптимізації продуктивності мережі. Аналізуючи мережні дані з різних джерел, включаючи пристрої, сервери та поведінку користувачів, алгоритми ML визначали точки перевантаження мережі та рекомендували оптимізовані стратегії маршрутизації [4].

Рисунок 1.6 – Алгоритми машинного навчання

Впровадження оптимізації продуктивності на основі ML призвело до значного зменшення затримки та покращення пропускної здатності мережі.

Адміністратори мереж змогли проактивно керувати мережним трафіком і ефективно розподіляти ресурси, забезпечуючи безперебійну роботу користувача навіть у періоди пікового використання.

1.4.2 Аналіз даних для виявлення несправностей

Багатонаціональна технологічна компанія впровадила методи аналізу даних для виявлення збоїв у мережі та усунення проблем у своїй глобальній мережній інфраструктурі [5]. Аналізуючи мережні дані в реальному часі, платформа аналізу даних виявляла аномалії та співвідносила їх із конкретними мережними пристроями та конфігураціями (рисунок 1.6). Цей підхід, орієнтований на ML, дозволив компанії швидко виявити основні причини збоїв у мережі та розпочати цілеспрямовані дії з усунення проблем. Впровадження спряло значному зниженню MTTR, що призвело до покращення доступності мережі та задоволеності клієнтів.

1.5 Майбутні тенденції та напрямки

Сфера машинного навчання та аналізу даних для мережних операцій продовжує розвиватися завдяки постійним досягненням і новим технологіям. Серед тенденцій та напрямків можна виділити деякі.

Розвиток методів і алгоритмів ML для мережних операцій. Дослідники постійно розробляють нові алгоритми ML, які можуть впоратися з унікальними проблемами мережних операцій. Ці алгоритми зосереджені на підвищенні точності прогнозування, масштабованості та універсальності в енергетичних мережних середовищах.

Інтеграція ML і аналізу даних із зростаючими інноваціями. ML і аналітика даних інтегруються з технологіями, що розвиваються, такими як 5G і периферійні обчислення [6]. Ця інтеграція дає змогу приймати рішення в реальному часі та організовувати оптимізацію на межі, гарантуючи низький

рівень бездіяльності та ефективний розподіл ресурсів.

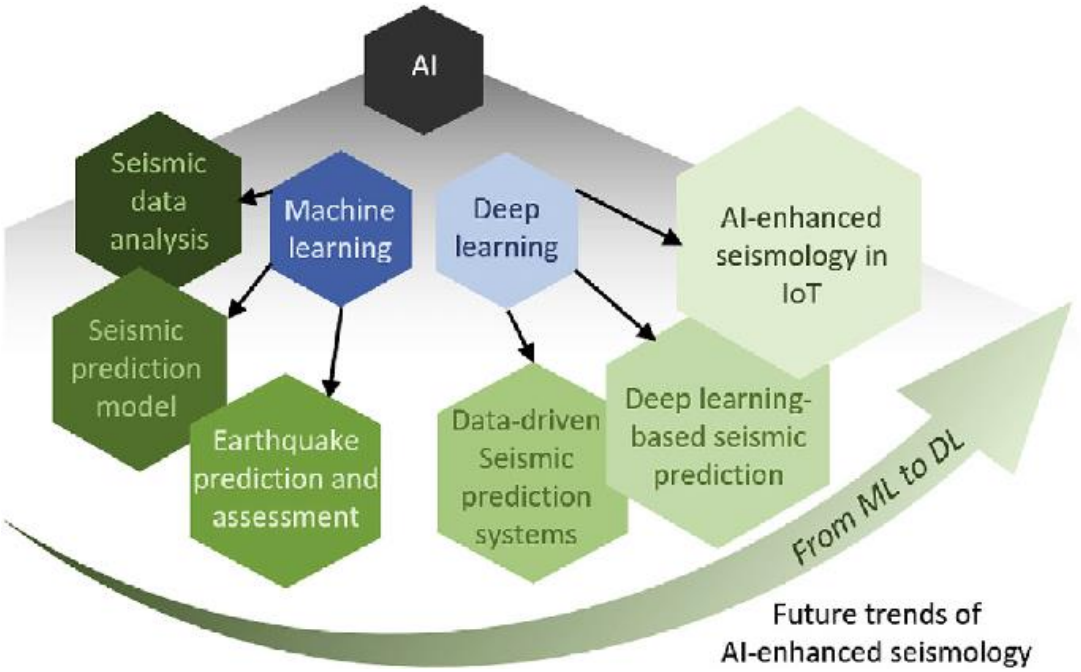


Рисунок 1.7 – Тенденції машинного навчання

Стандартизація та найкращі практики мережних операцій на основі ML. Оскільки ML і аналітика даних стають все більш поширеними в мережних операціях, розробка рамок стандартизації та передового досвіду стає вирішальною. Ці структури допоможуть керувати впровадженням, розгортанням і керуванням мережними операціями, керованим машинним навчанням, забезпечуючи достовірність.

Автономне керування мережею. Алгоритми машинного навчання будуть використовуватися для того, щоб зробити незалежні структури здатними керувати мережними операціями без втручання людини. Ці системи постійно відстежуватимуть роботу мережі, виявлятимуть особливості та, відповідно, коригуватимуть мережні конфігурації для покращення роботи та обмеження часу простою [6].

Прогнозне технічне обслуговування. Моделі машинного навчання використовуватимуться для передбачення апаратних збоїв і виконання

проактивної підтримки. Вивчаючи достовірну інформацію та показання датчиків у реальному часі, ці моделі можуть розрізняти дизайни та маркери потенційних розчарувань, дозволяючи адміністраторам мереж вирішувати проблеми до того, як вони спричинять критичні порушення.

Безпека мережі та виявлення загроз. Алгоритми машинного навчання будуть використовуватися для просування безпеки мережі та розрізнення нових загроз [7]. Ці обчислення можуть досліджувати організовану діяльність, виявляти аномалії та потенційні порушення безпеки чи зловмисну діяльність. Такий проактивний підхід покращить безпеку мережі та зменшить наслідки кібератак.

Оптимізація мережі та розподіл ресурсів. Алгоритми машинного навчання використовуватимуться для оновлення та організації розподілу активів, таких як швидкість передачі, контроль і обчислювальні ресурси. Ці моделі можуть аналізувати історичні дані та шаблони використання мережі для прогнозування потреб у ресурсах, забезпечуючи ефективний розподіл і масштабування ресурсів на основі коливань попиту [7].

Граничні обчислення та розподілене машинне навчання. З розвитком периферійних обчислень моделі машинного навчання розгортатимуться та навчатимуться на межі мережі [8]. Цей підхід до розподіленого машинного навчання зменшує затримку та вимоги до пропускної здатності, обробляючи дані та приймаючи рішення локально. Він також дає змогу в реальному часі досліджувати інформацію, створену крайовими пристроями, такими як пристрої Інтернету речей.

Зрозумілий штучний інтелект (англ., AI – Artificial Intelligence) для мережних операцій. У міру того, як моделі машинного навчання стають складнішими та поширеними в мережних операціях, зростатиме потреба в зрозумілому AI. Адміністраторам мереж знадобиться уявлення про те, як моделі приймають рішення та рекомендації, щоб забезпечити прозорість, підзвітність і відповідність нормативним вимогам.

2 МАШИННЕ НАВЧАННЯ В ЗАДАЧІ NTMA

2.1 Загальні відомості

Протягом останніх років проблемі моніторингу і аналізу мережного трафіка (NTMA) приділяється багато уваги як важливій темі досліджень у забезпеченні продуктивності мереж [9]. В якості загальних рішень в керуванні мережею методи NTMA були задіяні як промисловістю, так і науковими колами. Незважаючи на те, що були введені різні методи NTMA, нові мережні технології та парадигми ускладнили створення ефективних мереж. Нові мережі з тисячами вузлів, наприклад IoT, потребують регулярного моніторингу, щоб підтримувати їхню продуктивність. Різні цілі в мережі змушують адміністраторів мережі оцінювати мережу з точки зору, наприклад, проблем безпеки, підтримки вимог до якості обслуговування (QoS) і покращення споживання ресурсів, тощо [10]. Вказані цілі досягаються шляхом застосування методів NTMA, наприклад виявлення аномалій, класифікації мережного трафіку, керування несправностями та прогнозування трафіку.

Методи NTMA поділяються на дві основні групи: активні методи та пасивні методи [5]. Активні методи передбачають генерацію та введення пробного трафіку в мережу, щоб дізнатися про стан мережі. Точніше, дані тестового трафіку вводяться в мережу на основі запланованої вибірки, а потім вимірюватимуться різні показники продуктивності мережі. Приклади показників включають пропускну здатність мережі, коефіцієнт втрат пакетів, затримку та тремтіння (або зміну затримки). Оскільки методи активного моніторингу надають інформацію про ефективність у реальному часі, вони є основними методами контролю послуг на основі угоди про рівень обслуговування (англ., SLA – Service Level Agreement). Навпаки, пасивні методи в основному використовуються для моніторингу та аналізу реального

мережного трафіку в мережі. Пасивні методи викликають великий інтерес у промисловості для цілей управління та планування [11]. Для пасивних методів не потрібен інший сайт у мережі. Ці методи можна використовувати для ретельного моніторингу трафіку, особливо в ситуаціях після подій, наприклад, відмовостійкість і усунення несправностей. Крім того, вони ідеально підходять для отримання глибокого розуміння якості досвіду (англ., QoE – Quality of Experience) користувача. Застосування активних і пасивних методів узагальнено в таблиці 2.1.

Таблиця 2.1 – Категорії застосування активних та пасивних методів

Активні методи	Пасивні методи
Прямий і наскрізний аналіз	Комплексні траси для усунення несправностей
Якість обслуговування (QoS)	Якість досвіду (QoE)
Моніторинг в реальному часі	Діагностика проблем протоколу
Моніторинг роботи мережі та послуг	Моніторинг не в реальному часі
Моніторинг наскрізних транспортних процесів у режимі реального часу	Моніторинг обслуговування та досвіду клієнтів

Зростання комунікаційних систем і мереж з точки зору кількості користувачів і обсягу згенерованого трафіку ставить перед NTMA різні щоденні проблеми, зокрема [12]:

- зберігання та аналіз даних трафіку;
- використання даних трафіку для бізнес-цілей через отримання розуміння;
- інтеграція даних трафіку;
- перевірка даних трафіку;
- безпека даних трафіку;
- отримання даних трафіку.

Безпрецедентне збільшення кількості підключених вузлів і обсягу

даних збільшує складність мережі, що вимагає продовження досліджень для аналізу та моніторингу продуктивності мережі. Крім того, наявність великої та різнорідної кількості даних трафіку вимагає прийняття нових підходів для моніторингу та аналізу даних керування мережею. Через ці проблеми більшість робіт зосереджуються саме на одному аспекті NTMA, наприклад, виявленні аномалій, класифікації трафіку або QoS.

Серед проблем, згаданих вище, збір даних трафіку представляє величезні технічні труднощі в області NTMA, особливо для активних вимірювань, оскільки потрібно використовувати зонди для оцінки прогресу важливих параметрів мережі з часом. Зонди є одними з найефективніших методів отримання уявлень про наскрізну продуктивність, яку відчувають кінцеві користувачі. Активні та пасивні зонди є двома поширеними стратегіями, які можуть покращити продуктивність наскрізних вимірювань і визначити QoS і QoE шляхом доставки детальних даних трафіку [13]. Активний зонд намагається емулювати мережний трафік, а потім надсилати емульований трафік у мережі для вимірювання наскрізної продуктивності (наприклад, затримки). У порівнянні з активними зондами, пасивні зонди представляють окрему точку зору мережі. Пасивні зонди розміщуються на посиланнях у мережі, і вони запитують увесь трафік, який передається через з’єднання, яке контролюється.

Що стосується конкретних мережних сценаріїв і цілей збору даних трафіку (наприклад, класифікація трафіку та виявлення вторгнень), можна визначити різні вимоги до збору даних трафіку. Іншими словами, не потрібно отримувати всі доступні дані з мережі в завданні збору трафіку. Таким чином, мережні пакети зазвичай розглядаються як центральні цілі, які слід досліджувати в завданнях збору даних трафіку. Щоб відстежувати мережний трафік для оцінки його продуктивності, існує два основні методи, включаючи поверхневу перевірку пакетів (SPI) і глибоку перевірку пакетів (DPI). Перший стосується збору інформації із заголовків пакетів мережного трафіку, другий обробляє весь вміст пакета, включаючи дані користувача.

Зонди можуть використовувати обидва методи для збору інформації про мережу, але DPI має деякі недоліки, зокрема:

- аналіз даних користувачів може поставити під загрозу конфіденційність користувачів;
- обробка всього пакету вимагає більше часу та ресурсів порівняно з обробкою заголовка;
- у деяких типах мережного трафіку, наприклад у віртуальній приватній мережі (VPN) і зашифрованому мережному трафіку, DPI не можна використовувати.

Виходячи з проблем, згаданих вище для використання DPI, більшість зондів у нових методах NTMA використовують SPI. Слід зазначити, що одним із значних викликів у NTMA є отримання великої кількості надійних даних трафіку. Щоб впоратися з цією проблемою, протягом останніх років були запропоновані деякі інструменти, наприклад, [14], як архітектури збору даних. Однак адаптивний та ефективний підхід до збору даних, який можна було б повсюдно використовувати в неоднорідних і великомасштабних сучасних мережах, досі відсутній.

Найпоширенішим форматом даних для збору мережного трафіку є мережні пакети. Однак більшість методів збору мережних пакетів стикаються з проблемою втрати пакетів, особливо коли йдеться про велику кількість трафіку. Крім того, ці методи мають труднощі з високошвидкісними з’єднаннями та стають неефективними через їх низьку здатність.

Ще одним популярним механізмом є збір даних на основі потоку. Потокова мережа – це набір мережних пакетів з однаковими функціями, наприклад IP-адресою джерела/одержувача та портами джерела/одержувача. Порівняно з механізмами на основі пакетів, методи збору даних на основі потоків можуть зменшити кількість необхідних завдань для аналізу пакетів і забезпечити кращу продуктивність, особливо в гігабітних мережах. Тим не менш, фільтрація пакетів і потоків може серйозно ускладнити ці методи.

Сучасні мережні рішення знаходяться під тиском нового явища, відомого як великі дані. Цей факт ґрунтується на особливих характеристиках даних керування мережею, наприклад великому обсязі, високій швидкості, високій достовірності та високій різноманітності [15]. Дані керування мережею стосуються всіх даних, які відображають мережну ситуацію, в основному отриманих із заголовків пакетів (функція на рівні пакетів), наприклад, затримки пакетів, позначок часу та типу пакету. Технології NTMA можна вважати одними з основних споживачів великих даних. Крім того, це стає критичною галуззю дослідження в контексті аналітики великих даних через складність даних. Звичайні методи обробки даних для NTMA включають:

- математичні та статистичні методи (наприклад, регресія для аналізу часових рядів);
- алгоритми машинного навчання (ML) і підходи до обробки великих даних (наприклад, контрольоване навчання для виявлення вторгнень).

Технології NTMA повинні виконувати послідовність кроків для перетворення необроблених даних трафіку в корисну інформацію. Використання звичайних методів для аналізу великих даних стикається з багатьма викликами та проблемами, включаючи точність, високошвидкісну аналітику та ефективну обробку великих даних у режимі реального часу. Крім того, на основі нових парадигм, таких як Інтернет речей (IoT), велика кількість підключених пристроїв щодня створює величезний обсяг необроблених даних, і, отже, потрібні ефективніші методології для моніторингу та аналізу такої величезної кількості необроблених даних більш ефективним способом з точки зору часу й простору обробки.

Отже, в задачі NTMA методам ML приділено багато уваги. Техніки ML згруповані в чотири групи:

- контрольоване навчання;
- навчання з частковим контролем;
- неконтрольоване навчання;

- навчання з підкріпленням.

Серед різноманітних методів машинного навчання глибоке навчання (DL) є ключовим кроком для значного полегшення аналітики та виявлення знань у сфері великих даних [15]. DL використовується в багатьох галузях, включаючи комп'ютерний зір, охорону здоров'я, транспорт і розумне землеробство. Крім того, DL також привернув увагу технологічних компаній. Великі компанії, такі як Twitter, YouTube і Facebook, виробляють величезні обсяги даних щодня, і, отже, для них надзвичайно важливо обробляти ці великі дані. Алгоритми DL використовуються для аналізу отриманих даних і вилучення значущої інформації, оскільки традиційні методи обробки даних практично неможливі для обробки такої величезної кількості даних. В даній роботі розглянуто злиття двох технологій, тобто NTMA і глибокого навчання, однак не розглянуті всі можливі застосування NTMA через їх велику кількість. Досліджуються лише деякі ключові задачі.

2.2 Огляд NTMA

NTMA є технологією моніторингу мережного трафіку з належним ступенем деталізації (наприклад, на рівні пакетів) і дозволяє глибоко зрозуміти роботу та продуктивність мережі, а також поведінку користувачів [16]. У контексті комунікаційних систем і мереж NTMA відіграє вирішальну роль у знаходженні відповідей на питання:

- як працюють мережі, і моніторинг продуктивності мереж;
- як споживачі використовують ресурси та оптимізують використання ресурсів;
- як ефективно контролювати та керувати телекомунікаційними інфраструктурами для надання SLA.

У зв'язку зі стрімким зростанням кількості підключених пристроїв і обсягом даних трафіку для забезпечення стабільності та доступності систем зв'язку потрібні більш досконалі методи NTMA. Тому доцільним є розгляд

загальної структури NTMA, яка складається з п’яти блоків. Більшість існуючих дослідницьких робіт повністю або частково дотримуються схеми, заснованої на рисунку 2.1.

Першим кроком є чітке визначення цілей NTMA. Як згадувалося вище, типові цілі включають класифікацію трафіку, прогнозування трафіку, керування несправностями та безпеку мережі. Залежно від поставленої цілі, можливо, доведеться працювати над різними підцілями. Наприклад, якщо метою NTMA є класифікація мережного трафіку, підціллю може бути класифікація даних трафіку в різні класи на основі їхніх міток, наприклад трафік VPN і не-VPN або Firefox і Chrome. Другим кроком є збір даних керування мережею за допомогою пасивних або активних методів моніторингу. Через те, що ці два методи надають різні види стану мережі, їх можна використовувати разом, щоб скористатися перевагами обох методів.

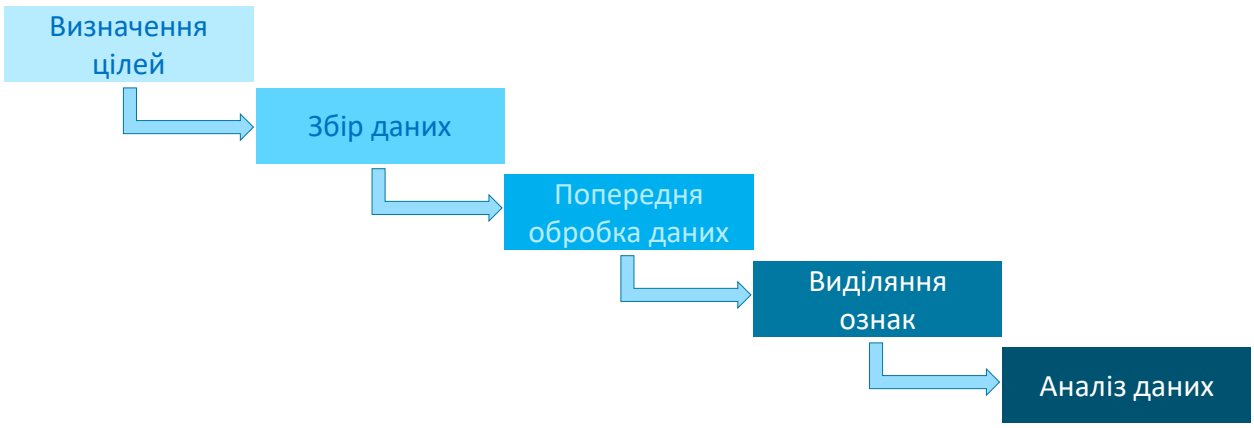


Рисунок 2.1 – Загальна структура процедури NTMA

На продуктивність NTMA, особливо підходів на основі ML, може значно вплинути попередня обробка та очищення даних. У комп’ютерній мережі розподіл функцій на рівні пакетів може змінюватися деякими звичайними діями, такими як повторна передача пакетів і дубльовані ACK. Отже, видалення таких даних керування мережею може покращити продуктивність застосунків NTMA, наприклад, прогнозування трафіку. Ще

одним методом попередньої обробки для покращення продуктивності програм NTMA є нормалізація. Особливо це надзвичайно важливо для підходів на основі ML та DL.

Після попередньої обробки даних NTMA реалізується етап вибору функцій, який дозволяє вибрати найбільш інформативні функції для досягнення мети. Вибір функції здійснюється автоматично або вручну. У першому випадку використовуються алгоритми вибору ознак для виділення найбільш релевантних ознак, а у другому – знання домену для виконання вибору ознак.

Після зазначених кроків експертами з аналізу даних виконується поглиблений аналіз попередньо оброблених даних для того, щоб отримати значущу інформацію. Як було зазначено вище, традиційними підходами для отримання значущих знань із необроблених даних є математичні та статистичні методи, алгоритми машинного навчання та підходи до великих даних. Для надійного та відтворюваного статистичного висновку важливим є вибір (з існуючих підходів) найбільш відповідної моделі чи методики. Підходи на основі машинного навчання мають перевагу над математичними та статистичними методами завдяки їхній здатності виявляти приховані закономірності в необроблених даних. На цей час існує чимало прикладів використання у комунікаційних системах і мережах методів ML у багатьох програмах, таких як системи виявлення вторгнень (IDS), виявлення аномалій, моніторингу, виявлення шаблонів.

Великі обсяги трафіку, який створюється людьми та машинами, наприклад, під час веб-серфінгу, вимагають розробки масштабованих алгоритмів та інструментів для обробки таких величезних обсягів даних за короткий проміжок часу. Задля цього застосовуються інфраструктури великих даних, такі як Hadoop і Spark [9]. Головним чином це зумовлено їхньою розподіленою архітектурою та можливістю прискорити процес за допомогою паралельної обробки та переміщення обчислювальної процедури до вузла, який генерує дані.

Слід зазначити, що дані керування мережею відрізняються від звичайних великих даних. Тому, щоб полегшити розуміння вимог до аналізу даних, вважається за необхідне дослідити характеристики даних керування мережею та висвітлити їхні основні відмінності від звичайних великих даних. Дані керування мережею мають кілька спільних характеристик, розглянутих далі.

Неоднорідність. Набір пристроїв, які обслуговуються в системі зв'язку та мережі, може бути дуже різнорідним, і ці пристрої споживають або генерують різні типи даних, що призводить до неоднорідності як мережного трафіку, так і даних керування мережею. Смартфони, транспортні засоби, датчики, інтелектуальна техніка та пристрої Інтернету речей є прикладами пристроїв, які можуть виграти від обслуговування систем зв'язку та мереж.

Кореляція часу та простору. Модель мережного трафіку, особливо стільникової мережі, демонструє дуже складну поведінку через різні фактори, зокрема мобільність і неоднорідність пристроїв, різні протоколи зв'язку, шаблони використання та вимоги користувачів. Крім того, в сучасних дослідженнях пропонується використовувати часові та просторові характеристики мережного трафіку та даних керування мережею з метою отримати більш детальне розуміння складної моделі, прихованої в даних мережного трафіку. Головним чином це пов'язано з тим, що багато програм і послуг надаються для певних місць, і, отже, часова та просторова інформація додається до даних трафіку та даних керування мережею.

Зашумлені дані. Під шумом розуміється будь-яка небажана зміна значення даних. У контексті мереж шум може створюватися деякими типовими подіями, спричиненими, наприклад, збоями, атаками тощо. Наприклад, у маршрутизації з кількома стрибками в мережах IoT неефективне керування чергами в середніх вузлах може спричинити тремтіння.

Високошвидкісні та великомасштабні потокові дані. Однією з відмінних характеристик даних мережного трафіку є потокове передавання

та висока швидкість передачі даних, особливо в таких службах, як потокове медіа, P2P-програми та потокова трансляція ігор. У цьому випадку дані керування мережею можуть залежати від обсягу та швидкості потоків.

Наслідки захисту даних. Поява протоколів шифрування мережного трафіку значно підвищила конфіденційність і безпеку спілкування. Використання технологій шифрування певною мірою гарантує, що сторонні особи не матимуть доступу до даних. Тим не менш, зростання популярності шифрування мережного трафіку ставить нові задачі перед NTMA. Наприклад, зашифрований трафік може знизити продуктивність IDS у виявленні шкідливого трафіку. Багато Інтернет-сервісів і програм використовують для безпечного зв'язку протоколи шифрування, наприклад захищений протокол передачі гіпертексту (HTTPS). Отже, в мережних пакетах залишається видимою невелика частка інформації, або доступною буде невідповідна інформація. Реалізація таких сценаріїв NTMA, як класифікація трафіку та керування несправностями, не є тривіальною. Наприклад, як метод класифікації, DPI [17] стикається з проблемами із зашифрованим трафіком і обмеженнями політики конфіденційності. Програми NTMA, що базуються на потоках, стикаються з меншими проблемами, пов'язаними з шифруванням, через те, що дані є послідовними, і потрібно менше передач. Однак вміст пакету стає більш незрозумілим, що робить аналіз мережного трафіку більш обмеженим. До того ж у деяких ситуаціях через шифрування на рівні IP приховуються заголовки TCP/UDP, стає майже неможливим дізнатися оригінальні номери портів.

Виділяння прихованих шаблонів і знань із великих даних є критичним завданням, однак не є занадто складним. Для такого вимогливого завдання, яке потребує здібностей, що виходять за межі звичайних механізмів навчання, потрібні нові підходи до навчання, моделі навчання та методи.

За останні кілька років увагу академічних кіл та промисловості привернули багато застосунків NTMA, наприклад класифікація трафіку, прогнозування трафіку та безпека мережі. Причина використання NTMA

полягає в тому, що програми NTMA відіграють важливу роль в керуванні мережею та ресурсами, методах аудиту мережі та виявленні вторгнень. Одним із потужних методів, що на основі штучного інтелекту надають розуміння комунікаційних систем і мереж, є глибоке навчання. Останніми роками DL використовується для багатьох програм NTMA, наприклад, для класифікації та прогнозування трафіку. Оскільки класичні алгоритми ML не здатні ефективно задовольняти нові аналітичні вимоги до комунікаційних систем і мереж, DL досягає все більшої популярності серед науковців.

Загалом, алгоритми DL представляють собою два значних удосконалення класичних методів машинного навчання. Вони усувають потребу у фазі розробки функцій через розгортання автоматичного навчання функцій. Таким чином, за допомогою алгоритмів DL можна легко отримати деякі корисні та значущі функції, які можуть бути неочевидними підходами при розробці вручну.

Алгоритми DL підвищують продуктивність навчання з точки зору точності та втрат через вивчення прихованих і високорівневих шаблонів із даних. Цього можна досягти шляхом передачі величезного обсягу даних трафіку в моделі DL.

2.3 Моделі глибокого навчання

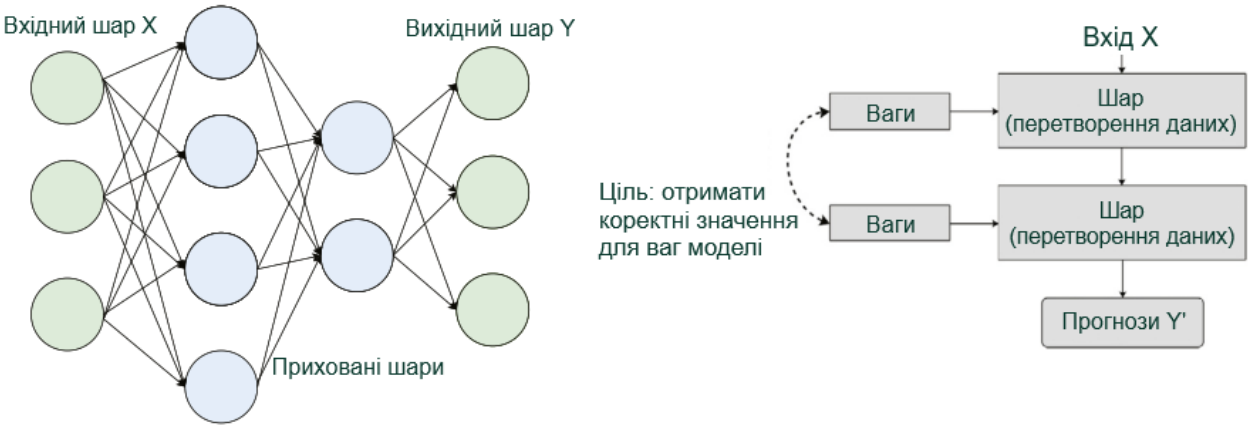
В останні роки штучний інтелект привернув великий інтерес до багатьох варіантів використання, таких як безпілотні автомобілі, чат-боти, віртуальні помічники тощо [18]. Історія штучного інтелекту починається з 1950-х років, коли дослідники намагалися автоматизувати інтелектуальні завдання, які зазвичай виконують люди. Довгий час багато експертів стверджували, що шляхом формулювання великого набору явних правил маніпулювання знаннями вони можуть реалізувати штучний інтелект, схожий на людину. Цей підхід, також відомий як символічний AI, був домінуючим методом досягнення штучного інтелекту людського рівня в

період з 1950-х до кінця 1980-х років. Незважаючи на те, що символічний AI успішно справлявся з чітко визначеними завданнями, такими як гра в шахи, він зіткнувся з труднощами при вирішенні більш складних завдань, таких як розпізнавання мови та класифікація зображень. Для вирішення цієї проблеми, в якості нового підходу було запропоноване машинне навчання.

Поява машинного навчання відкриває нову парадигму програмування. У парадигмі символічного AI людина-агент вводить правила (програму) і дані, якими потрібно маніпулювати відповідно до цих правил, і дає результати. Навпаки, у машинному навчанні агент-людина вводить дані та очікувані результати від даних, а потім модель навчання дає правила. Потім ці правила застосовуються до нових даних, щоб отримати оригінальні результати. Системи машинного навчання піддаються навчанню, а не явному програмуванню. Це означає надсилання величезної кількості даних у ці системи для пошуку значущих функцій у цих даних. Потім ці функції можна використовувати для створення правил для автоматизації завдання. Машинне навчання зазвичай має проблеми з великими та складними наборами даних, такими як набори даних зображень із тисячами або навіть мільйонами екземплярів. Для класичного статистичного аналізу, такого як байєсівський аналіз, практично неможливо працювати з такими великими наборами даних. Отже, машинне навчання, і особливо DL, демонструє відносно мало теорії математики та є інженерно-орієнтованим підходом.

DL – це спеціальне підполе ML, у якому для пошуку представлення даних на кожному рівні використовується глибока нейронна мережа (англ., DNN – Deep Neural Network) [19]. Глибина у визначенні DL відноситься до ідеї послідовних шарів представлень. З іншого боку, кількість шарів для моделювання даних відома як глибина моделі. Для таких складних завдань, як розпізнавання зображень, моделі DL часто мають десятки або навіть сотні послідовних шарів представлень. На відміну від DL, інші моделі машинного навчання часто включають один або два шари для представлення даних. Архітектура DNN представлена на рисунку 2.2 (а).

Як загальне визначення, можна стверджувати, що машинне навчання – це зіставлення вхідних даних (наприклад, відео та зображень) із цілями (наприклад, мітка «собака»), що досягається шляхом надання моделі багатьох екземплярів введення та цілей. Подібним чином видно, що DL виконує відображення введення в ціль через глибокі послідовні рівні перетворень даних. DL-модель вивчає ці перетворення, спостерігаючи за багатьма прикладами вхідних/цільових даних.



(а) (б)

Рисунок 2.2 – Архітектура мережі глибокого навчання та процес навчання

У DL-моделі ваги шару, також відомі як параметри, визначають, які перетворення будуть виконані з вхідними даними шару. Згідно з простим визначенням «ваги», це набір чисел (рисунок 2.2 (б)). У контексті DL під навчанням розуміється пошук набору правильних значень для ваг усіх шарів у моделі, щоб модель точно відображала вхідні дані для відповідних цілей. Через те, що моделі DL можуть мати десятки мільйонів параметрів (ваги), визначення правильного значення для всіх цих параметрів є складним завданням. На рисунку 2.3 спрощено показано зв’язок між AI, машинним навчанням і DL.

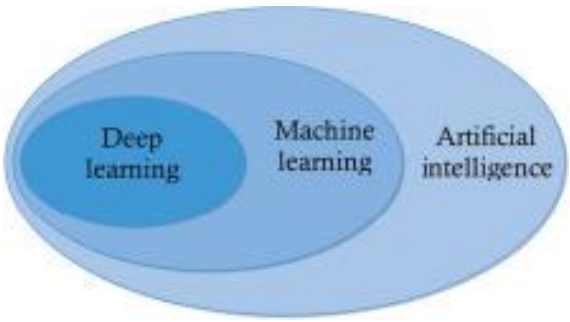


Рисунок 2.3 – AI, ML і DL

2.4 DL і NTMA

Методи машинного навчання, особливо алгоритми DL, є одними з найпопулярніших методів обробки даних мережного трафіку. Можливо, це пояснюється тим фактом, що сучасні комунікаційні системи та мережі, наприклад, IoT та стільникові мережі, мають особливі характеристики, які відповідають алгоритмам DL. Ці особливості включають генерацію великих даних, складність, мультимодальні дані, масштабність, зростаючу кількість протоколів у таких мережах тощо. Традиційні методи для NTMA мають свої проблеми; наприклад, вони є неточними або сильно залежать від людей-експертів. На відміну від традиційних методів, методи на основі DL мають деякі переваги для використання як методи NTMA, перелічені нижче.

Моделі DL не потребують значних людських зусиль і не залежать від вибору функцій. DL-моделі можуть використовувати різні репрезентативні шари та ефективні алгоритми для вилучення прихованих знань із величезних обсягів даних трафіку без розробки функцій. Ця перевага моделей DL дуже ефективна для методів NTMA, оскільки більшість даних керування мережею є немаркованими або напівмаркованими.

Моделі DL (наприклад, LSTM) здатні працювати з часово-просторовими даними, фіксуючи пов’язані залежності. Більшість даних про керування мережею, зібраних як набори даних часових рядів, можна аналізувати моделями DL з високою точністю. Розгортання точних і

ефективних методів для різних застосунків NTMA є надзвичайно важливим. Наприклад, точне передбачення мобільного трафіку є важливим для проектування трафіку (розподіл ресурсів на вимогу), економії енергії та аналізу мобільності користувачів у стільникових мережах (прогнозування руху).

У сучасних обчислювальних парадигмах, наприклад, Fog і Edge, задіяні пристрої оснащені високопродуктивним обчислювальним обладнанням, наприклад, графічним процесором (GPU) для обробки даних. Оскільки ці обчислювальні парадигми широко використовуються для виконання NTMA, методи DL можуть бути реалізовані, наприклад, обладнанням Fog і Edge для моніторингу мережі. Крім того, нові парадигми машинного навчання, наприклад федеративне навчання, в основному розроблені для реалізації методів глибокого навчання розподіленим способом [20]. Реалізація моделей DL за допомогою нових парадигм ML дозволяє DL навчати свою модель окремо на кожній машині. Це вважається великою перевагою, оскільки методи NTMA потребують збору інформації про керування мережею з різних машин до центральної точки. Використовуючи методи розподіленого машинного навчання, моделі DL можна навчати окремо на кожній машині, зменшуючи накладні витрати на мережу та загрозу безпеці та конфіденційності. На рисунку 2.4 показані основні типи застосувань NTMA.

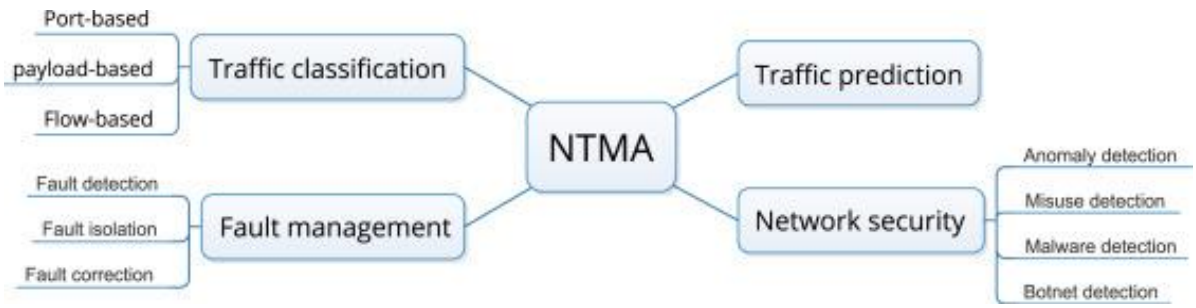


Рисунок 2.4 – Основні застосування NTMA

3 МОДЕЛЬ АНАЛІТИКИ МЕРЕЖНИХ ДАНИХ У СТІЛЬНИКОВИХ МЕРЕЖАХ 5G ІЗ ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ

3.1 Загальні відомості

Розвиток бездротових стільникових мереж пов’язаний зі значним збільшенням кількості користувачів і кінцевих точок, що призводить до створення дуже складних і завантажених мереж. Стандартизація систем стільникового зв’язку четвертого покоління (4G) у рамках Проекту партнерства третього покоління (3GPP) дозволила користувачам досягти пари сотень Мбіт/с, отримуючи доступ до застосунків, які потребують високої швидкості передачі даних, таких як телебачення високої роздільності. Проте 4G не в змозі задовольнити експоненціально зростаючі потреби користувачів. Зокрема, Cisco Visual Networking Index прогнозує, що кількість пристроїв, підключених до мобільних пристроїв, досягне 12,3 мільярда, а середні смартфони генеруватимуть 132 ГБ трафіку на рік [21]. Крім того, швидка поява застосунків, що вимагають зв’язку між машинами (M2M) (наприклад, Інтернет речей), принесла нові вимоги, які не задовольнялися попередніми стільниковими технологіями, розробленими для зв’язку «людина-людина» (H2H). Таким чином, за останнє десятиліття з’явилися стільникові мережі п’ятого покоління (5G). 3GPP вперше випустив специфікації для неавтономного (NSA) режиму доступу 5G, який базується на існуючій інфраструктурі 4G, таким чином підтримуючи взаємодію між існуючими стільниковими технологіями.

Дослідники очікують, що 5G буде значною зміною парадигми, а не поступовим прогресом 4G [22]. З цією метою 3GPP опублікував ще одну серію специфікацій для автономної (SA) 5G, яка представляє нове хмарне ядро 5G, незалежне від стільникової інфраструктури 4G. 5G SA забезпечує

значно спрощену мережу радіодоступу (англ., RAN – Radio Access Network) і архітектуру пристроїв, оскільки це цільовий варіант архітектури 5G, який полегшує ширший спектр випадків використання для нових типів пристроїв і моделей зв'язку пристроїв, таких як зв'язок M2M. Зокрема, 5G SA представляє розширене пакетне ядро з урахуванням нових випадків використання, включаючи IoT [23]. Таким чином, деякі з уже існуючих мережних функцій (NF) в архітектурі на основі послуг (SBA) застарілих стільникових мереж (тобто 4G або нижче) замінено новими NF в 5G SBA. Наприклад, замість функції політики та правил тарифікації (PCRF) у 4G 5G має функцію контролю політики (PCF). Подібним чином замість запису даних заряджання (CDR) 5G має функцію заряджання (CHF).

Одна з таких функцій пов'язана з аналізом даних. Аналітика даних стала життєво необхідною для 5G SA, оскільки вона розроблена для підтримки швидкості передачі даних, яка може досягати гігабіт. Функція аналізу мережних даних (NWDAF) є однією з нещодавно запропонованих функцій аналізу даних для мереж 5G, і вона надає аналіз мережі іншим NF [24]. Для цього NWDAF може використовувати будь-який алгоритм машинного навчання або штучного інтелекту на основі вимог (наприклад, часових обмежень) споживаючого NF. Відповідно до консорціуму 3GPP, очікується, що NWDAF матиме кілька можливостей з яких в даній роботі розглянуті такі:

- інформування про аномальну поведінку для групи користувальницького обладнання (UE);
- інформування про очікувану поведінку для групи UE;
- оцінка продуктивності мережного навантаження в області інтересу.

Вважається, що вибрані можливості мають вирішальне значення для стійкості мережі та QoS.

Ще до того, як 3GPP представив NWDAF для стільникових мереж 5G, моделі AI/ML часто використовувалися в бездротових мережах, а також у багатьох інших областях. Тим не менш, у зв'язку з потребою в наднадійному

зв’язку з малою затримкою і безпрецедентному трафіку даних, який експоненціально зростає, використання моделей AI/ML у стільникових мережах стало серйозною необхідністю. Тому 3GPP представив NWDAF для виконання цієї вимоги [24].

Одним із основних питань при використанні ML є підготовка наборів даних. Найкращим варіантом є використання набору даних, зібраних із фактичного налаштування мережі. Альтернативним підходом може бути використання Generative Adversarial Networks (GAN) [25]. Перший варіант в поточних умовах є неможливим. Для використання другого потрібно навчити GAN за допомогою зразка з реального набору даних, якого також не існує. Декілька робіт у літературі вивчають концепції ML, AI та DL для стільникових мереж 5G з урахуванням або без урахування NWDAF. Проте наразі ці роботи є незрілими. Крім того, навіть деякі розділи в специфікаціях 3GPP щодо NWDAF наразі залишаються порожніми через новизну NWDAF. Більше того, існуючі набори даних не відповідають вимогам сценарію NWDAF у цьому документі. Тому ми створюємо наш синтетичний набір даних на основі специфікацій 3GPP, як ми обговорюємо в Розділі V.

Отже, спершу потрібно створити загальнодоступний набір даних 5G [26] на основі специфікацій 3GPP для мереж 5G. Згенерований набір даних 5G включає топологію з фіксованою кількістю чарунок для фіксованої кількості категорій абонентів, де різні типи пристроїв, що мають різні моделі трафіку (наприклад, стільниковий телефон, транспортний засіб), можуть підключатися до мережі. Моделюється кожна комірка, використовуючи набір функцій, які отримуються з інших NF: байти, передані протягом часу моніторингу, список категорій, пов’язаних з абонентом, ідентифікатор персонального обладнання та інформація про зону мережі (тобто ідентифікатор RRU чарунки). Для того, щоб зробити синтетичний набір даних 5G більш реалістичним, додаються такі аномалії, як несподіване збільшення трафіку даних через певну чарунку.

3.2 Функція аналітики мережних даних

NWDAF – функція аналізу даних у стільникових мережах 5G, яка забезпечує аналіз мережі за запитом від інших NF. В якості джерела даних NWDAF може використовувати будь-яку іншу NF. Це означає, що існує двосторонній зв’язок між NWDAF і NF, як показано на рисунку 3.1. Показаний Nnwdaf представляє сервісний інтерфейс NWDAF, а Nnf – сервісний інтерфейс будь-якого NF (наприклад, Npcf представляє сервісний нтерфейс PCF). Як видно з рисунку 3.1, NWDAF може або надавати дані аналізу мережі іншим NF (тобто аналітичну інформацію), або NF можуть запитувати підписку від NWDAF на доставку даних (тобто підписку на події) за допомогою інтерфейсу Nnwdaf. Крім того, NWDAF отримує дані з інших NF за допомогою інтерфейсу Nnf.

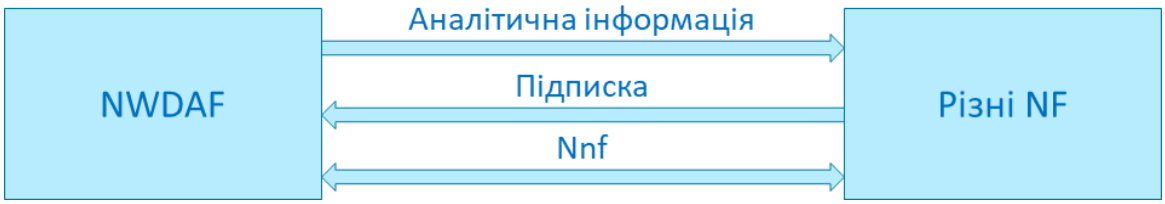


Рисунок 3.1 – Концептуальна модель збору і аналізу мережних даних

Є два сервіси Nnwdaf: сервіс підписки на події (тобто підписки на аналітику) та сервіс аналітичної інформації. Сервіс підписки на події Nnwdaf дозволяє споживачам NF підписуватися на різні аналітичні події та відмовлятися від них, а також сповіщає споживачів NF за допомогою відповідної підписки про спостережувані події. З іншого боку, сервіс аналітичної інформації Nnwdaf дає змогу споживачам NF запитувати й отримувати інший тип аналітичної інформації про подію від NWDAF. Таким чином, можна розглядати сервіс підписки на події більше як платну послугу, тоді як сервіс аналітичної інформації є таким, що потребує застосування

методів AI/ML. Враховуючи це, основна увага в роботі зосереджена на аналітичному інформаційному сервісі Nnwdaf. За допомогою аналітичної інформації Nnwdaf можна отримувати інформацію про таке:

- аномальна поведінка групи UE або конкретного UE;
- очікувана поведінка для групи UE або окремого UE;
- ефективність мережного навантаження в зоні інтересу;
- рівень навантаження екземпляра сегмента мережі;
- завантаження NF для конкретного NF;
- шаблон зв'язку для конкретного UE;
- перевантаження даних користувача в певному місці;
- мобільність групи UE або конкретного UE;
- статистика зміни якості обслуговування та потенційна зміна якості обслуговування в певній області;
- досвід обслуговування для програми або для сегмента мережі.

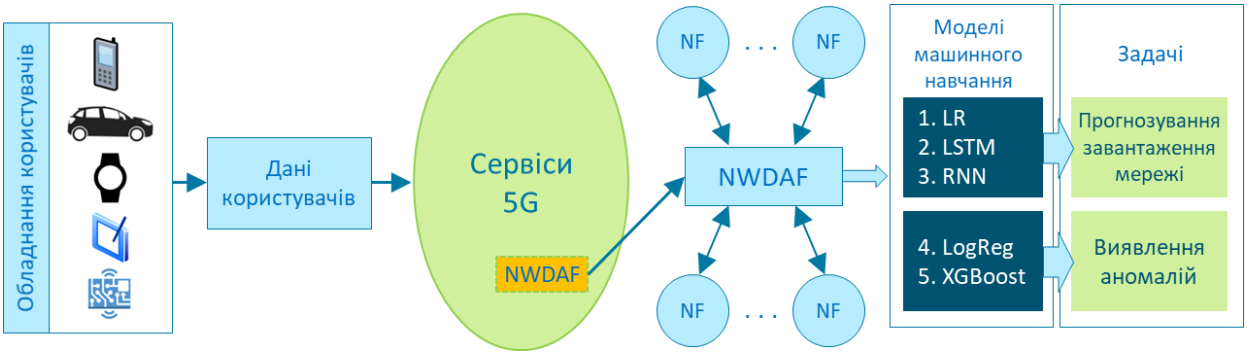


Рисунок 3.2 – Високорівневий робочий процес запропонованої системи

На рисунку 3.2 показано робочий процес, який складається з трьох етапів:

1) дані генеруються UE та надсилаються до 5G SBA;

2) NWDAF збирає цю інформацію (тобто інформацію, надіслану з UE) із пов'язаних NF;

3) функції ML у NWDAF виробляють аналітичну інформацію мережі.

У даній роботі поставлено акцент на перших трьох подіях, та класифікується інформація про поведінку для групи UE під час оцінки продуктивності мережного навантаження. Загальновідомих споживачів сервісу аналітичної інформації NWDAF можна побачити на рисунку 3.3. Будь-яка NF на рисунку 3.3 може бути споживачем подій, які аналізує NWDAF після підписки.

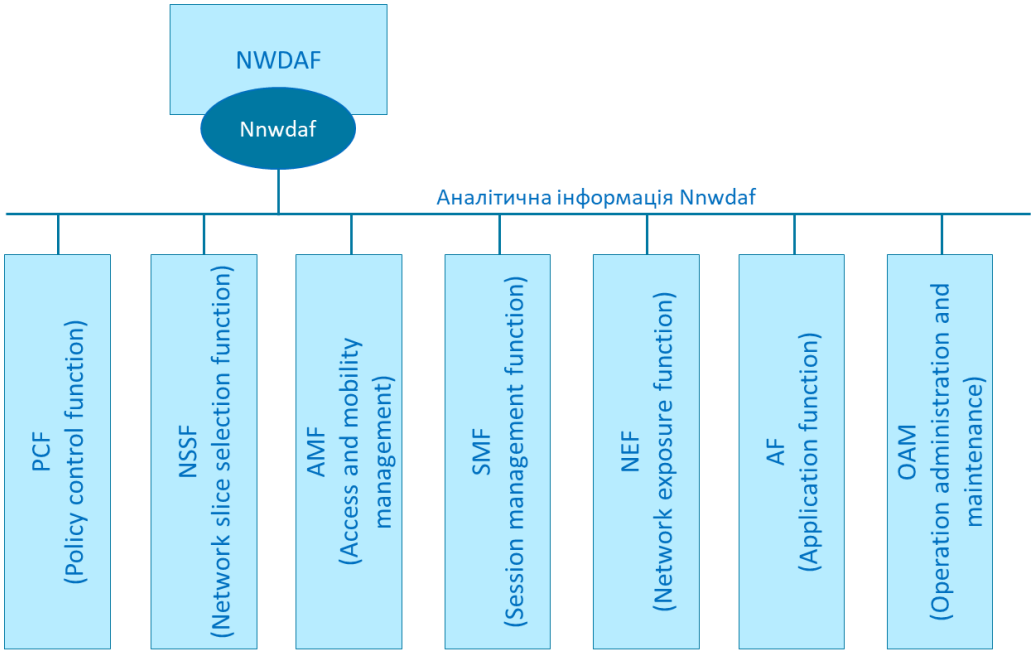


Рисунок 3.3 – Архітектура сервісу аналітичної інформації Nnwdaf

3.3 Модель системи

3.3.1 Робочий процес

Як показано на рисунку 3.2, дані, отримані від UE, передаються в 5G SBA, де знаходяться NWDAF та інші NF. NWDAF підключається до інших NF через інтерфейс на основі сервісів (SBI), і NWDAF та інші NF взаємно передають дані один одному. Потім NWDAF збирає інформацію з різних NF і адаптує кілька моделей ML для прогнозування продуктивності мережного

навантаження та виявлення його аномалій. Враховуючи фіксовану топологію, система використовує позначені дані для навчання; тому вибирається найкраща модель ML залежно від характеристики топології.

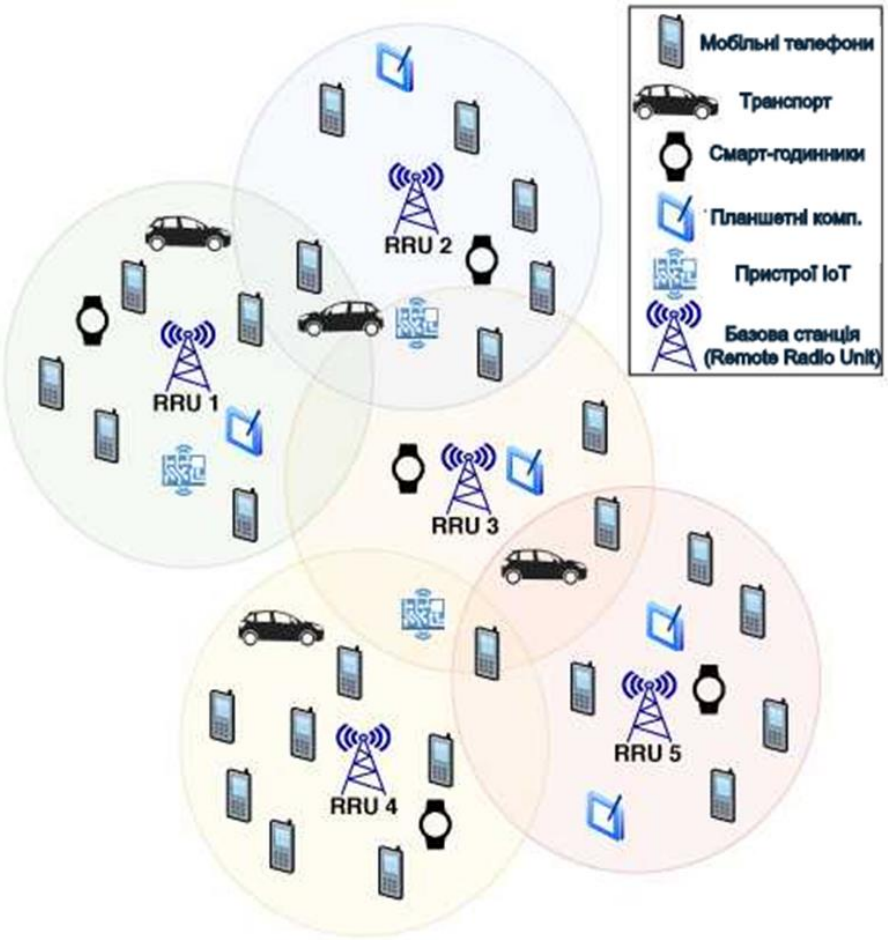


Рисунок 3.4 – Приклад представлення топології мережі

3.3.2 Топологія

Використано фіксовану стільникову топологію, яка складається з фіксованого набору чарунок RRU, фіксованого набору категорій абонентів і фіксованого набору типів персонального обладнання (тобто пристроїв). Заради простоти, незважаючи на те, що модель системи може підтримувати топології, які складаються з великої кількості комірок RRU, категорій абонентів і типів персонального обладнання, розглянуто топологію мережі, яка складається з п’яти чарунок RRU. У кожній із цих чарунок є три категорії

передплатників, де ці категорії передплатників представляють платинову, золоту та срібну підписки. Логіка цих трьох підписок полягає в тому, щоб зробити згенеровані дані більш реалістичними, оскільки постачальники послуг мобільного зв'язку продають такі види підписок у реальному світі. Крім того, у кожній категорії абонентів є п'ять різних типів персонального обладнання (тобто обладнання користувача), а саме: пристрій Інтернету речей, транспортний засіб, мобільний телефон, розумний годинник і планшетний комп'ютер. Зразок топології мережі можна побачити на рисунку 3.4.

3.3.3 Трафік

Згідно із запропонованою моделлю, кожне персональне обладнання в кожній категорії абонентів і чарунки RRU включає в себе заздалегідь визначений обсяг трафіку на початку моделювання. Тому можна сказати, що мережний трафік насичений від початку до кінця симуляції. Потім для кожного кроку часу моделювання (Δt) деяка частина навантаження переходить від однієї чарунки (тобто джерела) до іншої (тобто цільової), яка є суміжною з чарункою джерела. Процес передачі (хендоверу) також відбувається на основі заздалегідь визначених співвідношень. Щоб зробити трафік більш реалістичним, середні коефіцієнти хендоверу змінюються відповідно до часу доби (таблиця 3.2).

Між 22:00-06:00, оскільки очікується, що люди будуть менше рухатися, середній коефіцієнт хендоверу також буде меншим. З 06:00 до 07:00 та з 20:00 до 22:00, оскільки дороги, швидше за все, будуть вільними від руху, очікується, що люди рухатимуться швидше з одного місця в інше, а середній коефіцієнт хендоверу буде вищим. Між 11:00-13:00 трафік буде дещо більшим, і люди рухатимуться трохи повільніше порівняно з 06:00-07:00 та 20:00-22:00, і, отже, середній коефіцієнт хендоверу буде меншим. Між 09:30-11:00 і 16:00-20:00 середній коефіцієнт хендоверу очікується меншим, ніж

11:00-13:00 через збільшення трафіку. Оскільки 07:00-09:30 і 16:00-20:00 є годинами пік, очікується, що середній коефіцієнт хендоверу буде найменшим, за винятком нічного часу. Крім того, для пристроїв Інтернету речей не очікується істотної різниці в середньому коефіцієнті хендоверу для часу доби, оскільки очікується, що ці пристрої не рухатимуться так сильно, як інше персональне обладнання, яке тут розглядається. З іншого боку, середні коефіцієнти хендоверу вищі для транспортних засобів через їхню мобільність.

Таблиця 3.2 – Середні значення коефіцієнтів хендоверу за годину (%)

Часовий проміжок	Пристрій IoT	Транспорт-ний засіб	Мобільний телефон	Розумний годинник	Планшетний комп’ютер
00:00-06:00	1	10	2,5	2,5	1
06:00-07:00	1	18	4,5	4,5	1,8
07:00-09:30	1	12	3	3	1
09:30-11:00	1	14	3,5	3,5	1,2
11:00-13:00	1	16	4	4	1,5
13:00-16:00	1	14	3,5	3,5	1,2
16:00-20:00	1	12	3	3	1
20:00-22:00	1	18	4,5	4,5	1,8
22:00-00:00	1	10	2,5	2,5	1

Слід зауважити, що значення коефіцієнта хендоверу в таблиці 3.2 є середніми значеннями, які означають, що існують також значення дисперсії. За допомогою цих ретельно налаштованих статистичних параметрів за допомогою канонічної апрксимації можна наблизитись до реального трафіку користувачів.

3.4 Генерація даних

В історії інформатики існують різні типи моделей AI/ML. Серед багатьох моделей AI/ML алгоритми, що лежать в основі цих моделей, працюють по-різному. Оскільки в роботі розглядаються алгоритми контрольованого машинного навчання, в цьому контексті стають вирішальними дані з мітками. Враховуючи це, створено позначений набір даних для стільникових мереж 5G. Вибір полів запропонованого набору даних враховує специфікації 5G, опубліковані консорціумом 3GPP. Вибрані поля:

- швидкість передачі даних – кількість переданих даних у байтах за певний період часу;
- інформація про зону мережі – інформація про комірку групи UE, до якої підключено;
- категорії підписки – політикою групи UE є підписка.
- ідентифікатор персонального обладнання – інформація про тип пристрою UE (наприклад, мобільний телефон, розумний годинник).

Під час генерації позначеного набору даних враховуються наступні попередньо визначені параметри як поля введення моделювання генерації даних:

- номер комірки RRU, який представляє кількість комірок RRU у топології;
- назви та ідентифікатори категорій;
- ідентифікатори персонального обладнання та назви типів пристроїв;
- конфігурації початкового навантаження, які представляють початкове навантаження на групу особистого обладнання;
- конфігурації комірок суміжності;
- відсоток здачі за групу особистого спорядження;
- середнє значення та коефіцієнти дисперсії для процесу передачі;
- крок моделювання, який представляє період процесу отримання

даних з NF;

- час моделювання, який представляє загальну тривалість моделювання.

Для моделювання генерації даних, залежно від категорії абонента та типу персонального обладнання, визначені початкові конфігурації навантаження для кожної комірки RRU (таблиця 3.3). Як видно, різні навантаження призначені для кожного типу персонального обладнання, а також для кожного категорії абонентів. Логіка, що стоїть за цими значеннями навантаження, полягає в тому, що користувач, швидше за все, підпишеться на платинову категорію, а не на золоту та срібну категорії для підписки на мобільний телефон. З іншого боку, менша ймовірність підписки на вищу категорію для планшетного комп'ютера, транспортного засобу та пристрою IoT. Крім того, для комірки RRU з чотирма суміжними комірками RRU на рисунку 3.4 середні коефіцієнти хендоверу в таблиці 3.2 подвоюються, щоб зберегти баланс у мережі. Крім того, як зазначалося вище, коефіцієнти у таблиці 3.2 є середніми значеннями. Припустимо, що події хендовера демонструють гаусівський розподіл, а середнє значення та параметри дисперсії подано наступним чином:

$$\Delta H_{ratio} \sim N(\mu, \sigma^2), \tag{2.1}$$

де ΔH_{ratio} – результуючий коефіцієнт передачі, а $N(\mu, \sigma^2)$ – випадкова величина Гауса із середнім μ дисперсією σ^2.

Далі генеруться дані про мережний трафік за шість місяців, які складаються зі знімка мережі через кожний 15-хвилинний інтервал ($\triangle t$). У кожному з цих інтервалів UE може здійснювати передачу обслуговування між сусідніми чарунками.

Щоб зробити набір даних більш реалістичним, додаються аномалії до згенерованих даних мережного трафіку протягом усього моделювання.

Аномалії виглядають як несподівано створений великий обсяг мережного трафіку порівняно із середнім мережним трафіком, який з часом зменшується та стабілізується. Створюючи подібні аномалії, можна знайти паралелі із повсякденним життям, де постійно якісь відео стають вірусними або з'являються екстрені новини, що все більше впливає на дані мережного трафіку.

Таблиця 2.3 – Початкові навантаження (Гбіт/с)

Категорія абонента	Пристрій IoT	Транспорт-ний засіб	Мобільний телефон	Розумний годинник	Планшетний комп'ютер
Платина	3	20	90	1	6
Золото	4	18	72	1	5
Срібло	5	16	53	1	5

В описі аномалій також позначено моменти часу з аномальним навантаженням трафіку. Це потрібно для досягнення базової достовірності та отримання інформації про поведінку з набору даних мережі 5G.

Реалізовано методологію генерації даних мовою програмування Python. Це пов'язано з тим, що багато алгоритмів ML реалізовано у вигляді Python-бібліотек, оскільки ця мова широко використовується, її легко читати та кодувати. Окрім реалізованих бібліотек ML, для Python також доступні бібліотеки обробки даних, що робить цю мову програмування дуже зручною для досягнення поставлених цілей.

У реалізації створюються моделі для типу пристрою, категорії абонента та чарунки RRU. Об'єктна модель чарунки містить інформацію про суміжні чарунки, як зазначено раніше. Об'єктна модель категорії абонента зберігає інформацію для кожного типу пристрою, включаючи навантаження та статистику. Об'єктна модель типу пристрою обробляє операції хендоверу та зберігає інформацію про завантаження в поточний момент часу. Після визначення всіх трьох моделей для процесу генерації даних створюється

запропонована модель системи, використовуючи попередньо визначену матрицю суміжності для комірок RRU, параметри передачі для типів пристроїв, тривалість моделювання та відсоток випадків аномалії. Після визначення всього цього можна розпочинати моделювання створення даних. У кожну точку часу t (тобто 0, $\triangle$t, 2x$\triangle$t, …, t) кожен тип пристрою визначає, скільки навантаження буде передано для хендоверу, тоді операція хендоверу відбувається шляхом передачі мережного навантаження для конкретного типу пристрою від вхідної комірки RRU до цільової комірки RRU. Крім того, випадковим чином перед початком моделювання генерації даних вибирається час початку та закінчення аномалій. Під час моделювання, коли починається період аномалії, до кожного пристрою додається попередньо визначений відсоток навантаження мережі, яке експоненціально зростає протягом періоду аномалії. Після завершення процесу генерації даних експортуються усі дані про навантаження мережі для аналізу та створення відповідних функцій. На рисунку 3.5 представлено агреговану швидкість передачі даних для кожної комірки.

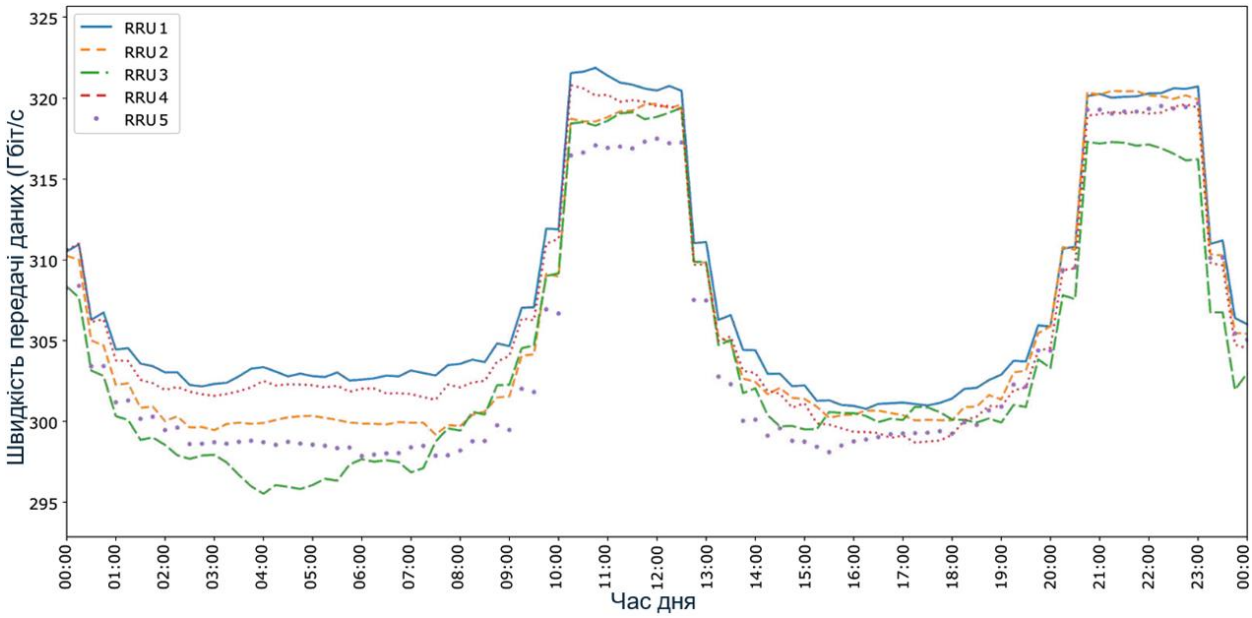


Рисунок 3.5 – Агрегована швидкість передачі даних на комірку RRU для вибіркового дня

3.5 Моделі машинного навчання

Моделі ML використовуються для отримання певної необхідної інформації з минулих даних. Існують різні типи алгоритмів машинного навчання, кожен із яких призначений для вирішення певного типу проблеми. Як згадувалося раніше, моделі ML, які використовуються в цій роботі, належать до категорії алгоритмів контрольованого навчання. Оскільки згенерований набір даних 5G є правильно позначеним, можна скористатися перевагами алгоритмів контрольованого навчання. В роботі розглядаються дві проблеми, для яких використовуються моделі ML для різних рішень і порівнюється продуктивність цих моделей.

3.5.1 Виділяння ознак

Для того, щоб отримувати точні результати, моделі ML потрібно навчити. Дані повинні бути оброблені та приведені у форму, зрозумілу моделі ML. Щоб навчити моделі ML, деякі з певних функцій потрібно витягти з набору даних. Завдяки людському розумінню та детальному аналізу даних ці функції стають дуже корисними для алгоритму ML протягом усього процесу навчання, що призводить до кращих результатів прогнозування. З іншого боку, DL-моделі, як правило, здатні вивчати шаблони даних без допомоги цих ідей і функцій через свою нейромережну природу. Проте для справедливого порівняння слід подавати однакові дані для моделей ML і DL.

У процесі виділяння ознак створюються такі ознаки, які враховують швидкість передачі даних у попередніх часових інтервалах. Виділені ознаки, зокрема, такі:

- last2_mean – середня швидкість передачі даних за останні два $\triangle t$;
- last4_mean – середня швидкість передачі даних за останні чотири $\triangle t$;
- last8_mean – середня швидкість передачі даних за останні вісім $\triangle t$;

- per_change_last2 – відсоток різниці швидкості передачі даних між останніми двома △t;
- per_change_last3 – відсоток різниці швидкості передачі даних між t-△t та t-3x△t;
- per_change_last4 – відсоток різниці швидкості передачі даних між t-△t та t-4x△t;
- change_last2 – різниця в швидкості передачі даних між останніми двома △t;
- change_last3 – різниця швидкості передачі даних між t-△t та t-3x△t;
- change_last4 – різниця швидкості передачі даних між t-△t та t-4x△t.

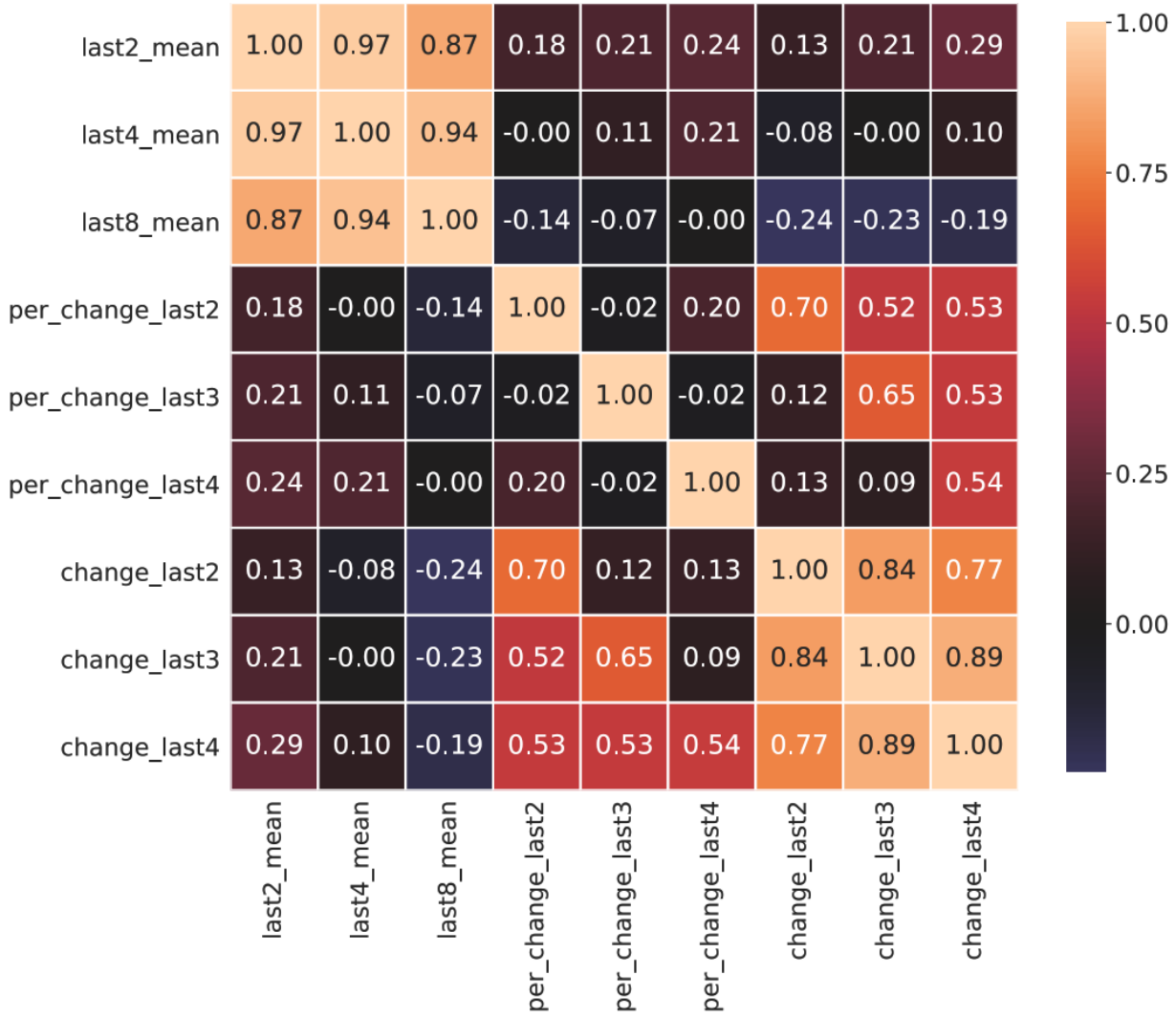


Рисунок 3.6 – Кореляційна матриця виділених ознак

Серед усіх розглянутих ознак виконуються тести на важливість ознак і кореляцію. Ці два тести дають інформацію про якість функцій, що впливає на якість навчених моделей. Тест важливості ознак показує, що last2_mean, last4_mean і last8_mean є найважливішими ознаками. За цими ознаками йдуть per_change_last2, per_change_last3 і per_change_last4, за якими слідують per_change_last2, per_change_last3 і per_change_last4. Однак, як видно на рисунку 3.6, середні значення швидкості передачі даних і зміна швидкості передачі даних мають вищу кореляцію між їхньою підгрупою ознак порівняно з відсотковою зміною ознак швидкості передачі даних. Зважаючи на перевірку важливості ознак, вибрано найважливішу ознаку з характеристик середньої швидкості передачі даних, якою є last2_mean і всі відсоткові зміни ознак швидкості передачі даних. Ознаки зміни швидкості передачі даних не вибрано через їх низькі бали порівняно з іншими в тесті важливості ознак.

3.5.2 Прогноз продуктивності навантаження мережі

Як було показано раніше, однією з головних задач цієї роботи є оцінка ефективності навантаження мережі. Цю проблему можна визначити як проблему часових рядів, а потім використовувати три різні моделі ML, а саме LR, RNN і LSTM, де LSTM є дещо модифікованою версією RNN. Під час навчання цих моделей ML використовується позначений набір даних (тобто міткою є швидкість передачі даних кожного часового інтервалу) і виділені ознаки (3.5.1).

Лінійна регресія. LR відповідає лінійному відношенню між даними характеристиками багатьох спостережень і міток. Оскільки LR є однією з найбільш часто використовуваних моделей ML для передбачень і проблем прогнозування, цю модель вибрано для порівняння як базову.

Рекурсивні нейронні мережі (англ., RNN – Recursive Neural Network). Нейронні мережі та глибоке навчання забезпечують ефективне вирішення

багатьох проблем. Природа нейронних мереж дозволяє алгоритму вивчати складні зв’язки в функціях і виробляти високоточні результати. RNN – це спеціалізована версія алгоритму нейронної мережі, що дозволяє переносити дані з попередніх нейронів і підвищує точність короткострокового та довгострокового прогнозування. В цій роботі RNN використовується для короткострокового прогнозування навантаження на мережу. Модель складається з одного простого шару RNN, чотирьох прихованих шарів і одного вихідного шару. Функція втрат RNN – це середня абсолютна похибка (MAE), яка також є одним із показників продуктивності.

Довга короткочасна пам’ять (англ., LSTM –Long Short-Term Memory) є модифікованою версією RNN, де структура нейронів змінена порівняно з RNN. У той час як RNN зберігає інформацію з попередніх нейронів, складна структура LSTM допомагає зберігати інформацію з дуже давніх даних на відміну від простих RNN. Таким чином, навіть якщо між спостережуваною картиною є часовий проміжок, LSTM може робити точні прогнози. У моделі LSTM використано один шар LSTM, чотири приховані шари та один вихідний шар. Функція втрат LSTM також є MAE, подібною до моделі RNN.

3.5.3 Виявлення аномалій

Для отримання інформації про поведінку групи UE, під час створення набору даних мережного трафіку додаються аномалії навантаження мережі, а також позначається кожен часовий інтервал як нормальний або ненормальний. Ця інформація про поведінку класифікується за допомогою двох різних моделей ML, а саме логістичної регресії та екстремального градієнтного підсилювання.

Логістична регресія – широко використовувана модель для різних проблем класифікації. Логістична регресія використовує логістичну функцію для оцінки міток заданих даних. Логістична регресія вибрана в якості базової моделі для проблеми виявлення аномалії. Оскільки отримані дані мають

неоднакову кількість аномальних і нормальних станів протягом усього моделювання, потрібно встановити параметри ваги класу, щоб подолати проблеми, які виникнуть через незбалансовану кількість типів міток (тобто аномалія та норма).

Екстремальне підсилювання градієнта. Для проблем класифікації існує багато алгоритмів, які використовують підходи на основі дерева. Алгоритми на основі дерева приймають рішення на основі заданих характеристик, одночасно зменшуючи функцію втрат. XGBoost – це найсучасніша реалізація дерев рішень із посиленням градієнта, призначена для швидкості та продуктивності. Це широко використовуваний алгоритм для задач класифікації, отже модель XGBoost використано у цій роботі. Як зазначалося вище, отримані дані мають незбалансовану кількість аномалій. Щоб усунути проблеми, спричинені незбалансованою кількістю міток, також слід налаштувати модель XGBoost, зменшуючи ефективність домінуючої мітки. Ця методологія допомагає моделі отримати вищий бал з точки зору метрики ефективності AUC-ROC.

4 ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

Враховуючи сценарії, описані в розділі 3, і дані, отримані відповідно до розглянутих процедур, отримано результати, які розглядаються у двох частинах:

- дослідження продуктивності мережного навантаження за допомогою моделей LR, LSTM і RNN;
- дослідження аномалій в мережі протягом моделювання за допомогою моделей логістичної регресії та XGBoost.

4.1 Прогноз продуктивності мережного навантаження

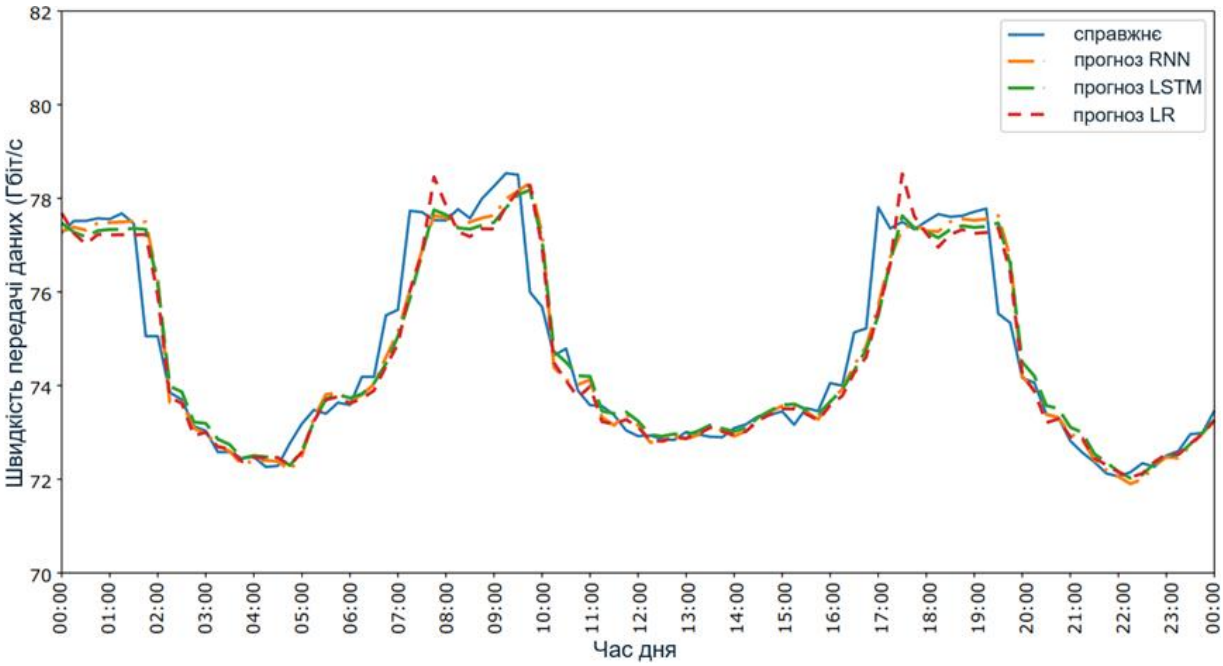
Використовуючи згадані вище три моделі AI/ML, виконано моделювання для кожного обладнання користувача в кожній категорії абонентів і чарунки RRU. Потім обчислено середню відсоткову помилку (MAPE) і середню абсолютну помилку (MAE) для кожного з цих сценаріїв. Результати з використанням цих показників можна побачити в таблиці 4.1. Слід зауважити, що Cell представляє номер комірки RRU, як показано на рис. 3.4, а SubsCat представляє категорії абонентів (платина, золото та срібло – від 1 до 3, відповідно). У таблиці 4.1 можна побачити, що LSTM і RNN працюють краще, ніж LR, у всіх сценаріях, як і очікувалося. Майже в половині сценаріїв RNN перевершує LSTM. Причина такої конкуренції зумовлена природою нейронних мереж, яка включає значний фактор рандомізації. Однак у середньому LSTM також перевершує RNN за показником MAPE. Проте, як видно, RNN перевершує LSTM за показником MAE. Це пояснюється тим, що RNN успішніше виявляє несподівані умови (тобто нестабільні швидкості передачі даних). Іншими словами, LSTM успішніше виявляє постійні швидкості передачі даних порівняно з RNN. Незважаючи на те, що оцінка RNN за метрикою MAE нижча, оскільки

непостійні швидкості передачі даних вищі за постійні, за метрикою MAPE оцінка LSTM нижча порівняно з RNN через співвідношення між чисельником і знаменником.

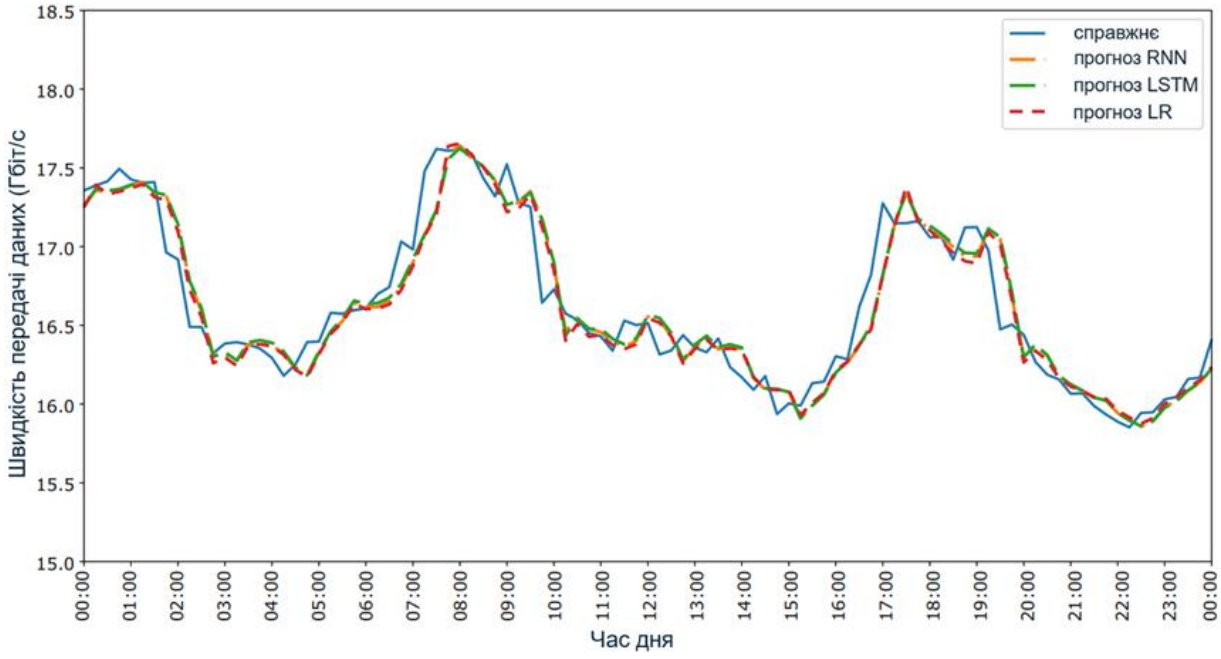
Таблиця 4.1 – Результати для прогнозування продуктивності навантаження

(Cell - SubsCat) ID	MAPE			MAE		
	LR	LSTM	RNN	LR	LSTM	LSTM
1 - 1	0,577	0,504	0,512	0,185	0,157	0,148
1 - 2	0,575	0,512	0,573	0,237	0,205	0,233
1 - 3	0,579	0,511	0,498	0,290	0,251	0,217
2 - 1	0,578	0,510	0,499	0,184	0,158	0,139
2 - 2	0,576	0,510	0,521	0,238	0,202	0,192
2 - 3	0,585	0,504	0,519	0,291	0,242	0,219
3 - 1	0,761	0,680	0,735	0,223	0,196	0,204
3 - 2	0,757	0,688	0,754	0,282	0,252	0,269
3 - 3	0,750	0,696	0,735	0,339	0,312	0,307
4 - 1	0,581	0,507	0,487	0,185	0,157	0,135
4 - 2	0,576	0,505	0,501	0,238	0,205	0,180
4 - 3	0,581	0,505	0,499	0,289	0,243	0,218
5 - 1	0,578	0,506	0,515	0,185	0,159	0,147
5 - 2	0,581	0,511	0,539	0,238	0,204	0,197
5 - 3	0,583	0,509	0,500	0,289	0,243	0,214
Середнє	0,615	0,544	0,560	0,247	0,213	0,202

На рисунку 4.1 можна побачити прогнози швидкості передачі даних моделей RNN, LSTM і LR, а також фактичні результати протягом усього часу, які випадковим чином фіксуються за один повний день у шестимісячному наборі даних. Із залежностей, показаних на рисунку 4.1, можна дійти до двох чітких висновків.



(а) номер комірки RRU – три, категорія абонента – золото, тип UE – стільниковий телефон



(б) номер комірки RRU – чотири, категорія абонента – платинова, тип UE – транспортний засіб

Рисунок 4.1 – Залежність швидкості передачі даних від часу для різних моделей AI/ML

По-перше, моделі ШНМ перевершують LR, показуючи, що вони краще передбачають раптові зміни (тобто різкі схили). Оскільки LR не надає складної формули для прогнозів, ці результати були очікуваними. По-друге, для постійних швидкостей передачі даних можна побачити, що прогнози LSTM ближчі до фактичних значень, за якими слідують прогнози RNN і пізніше прогнози LR. Однак для непостійних швидкостей передачі даних прогнози RNN ближчі до фактичних значень, за якими слідують прогнози LSTM, які також супроводжуються прогнозами LR. Можна чітко побачити ці два спостереження, подивившись між 07:00 і 10:00 для нестабільної швидкості передачі даних на рисунку 4.1(а), і між 00:00 і 01:30 на рисунку 4.1(б).

І нарешті, хоча LR працює прийнятно із синтетично згенерованим набором даних, слід зазначити, що регресійні моделі можуть працювати гірше, а отже, ймовірно, забезпечуватимуть нижчі значення точності в фактичних розгортаннях. Набір даних створено із високим стандартним відхиленням та параметрами рандомізації, щоб зробити його максимально реалістичним. Незважаючи на те, що базова ідея навчання моделей ML однакова, можна очікувати, що результати в базових моделях ML будуть дещо іншими. З іншого боку, очікується, що складні моделі ML забезпечуватимуть подібну продуктивність порівняно з продуктивністю запропонованого набору даних.

4.2 Виявлення аномалій

Для виявлення аномалій у даних мережного трафіку 5G використовується логістична регресія та моделі XGBoost. Щоб виміряти продуктивність цих двох моделей, використано оцінку AUC-ROC. Крім того, результати візуалізуються за допомогою кривих робочих характеристик приймача (ROC). Крива ROC має дві осі, а саме частоту справжніх позитивних сигналів (тобто ймовірність виявлення P_d) і частоту помилкових

позитивних сигналів (тобто частоту помилкових тривог P_f). P_d – це відношення справжніх позитивних результатів до фактичних справжніх значень. Цей показник представляє точність моделі для виявлення аномалій (тобто відсоток успішних прогнозів серед станів мережі, які мають аномалію). Інша вісь, P_f, є відношенням хибних позитивних результатів до фактичних негативних значень. Цей показник представляє ймовірність помилкових рішень, коли фактичний стан мережі не демонструє аномалій (тобто відсоток помилкових прогнозів для станів без аномалій). Бажана форма кривої ROC схожа на лікоть, що йде до верхньої сторони фігури.

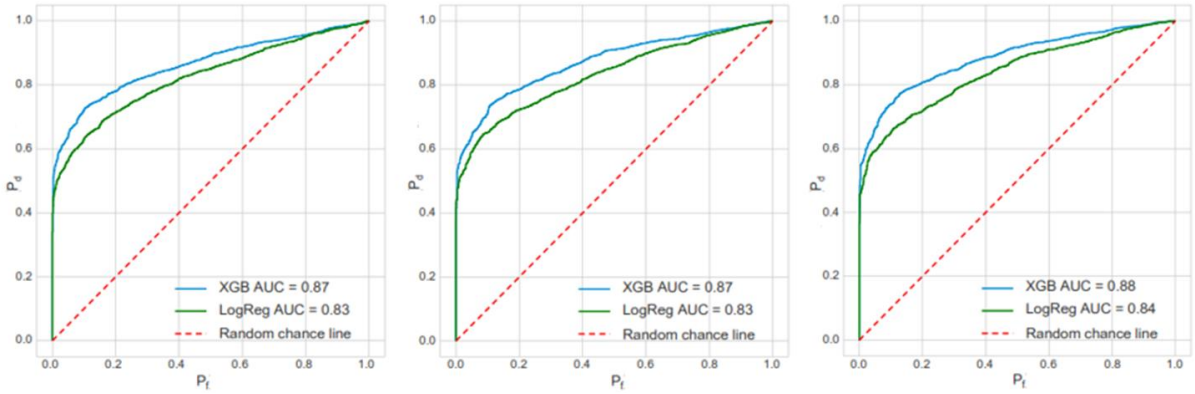
Показники AUC-ROC, точності та достовірності підсумовано для моделей XGBoost і логістичної регресії в таблиці 4.2. Як видно з таблиці 4.2, показник AUC-ROC моделі XGBoost значно вищий, ніж модель логістичної регресії для кожної категорії абонентів і чарунки RRU. Як також видно з середнього показника AUC-ROC, точності та достовірності, модель XG-Boost прогнозує аномалії набагато краще, ніж модель логістичної регресії. Хоча точність значно не покращилася, спостерігаємо значне покращення достовірності.

На рисунку 4.2 порівнюються результати різних категорій абонентів зі стільниковими телефонами, тобто під номерами три і чотири RRU. Порівнюючи залежності на рисунку 4.2, можна побачити, що різниця площі під кривою ROC між XGBoost і логістичною регресією вища на користь моделі XGBoost. Загалом, серед чарунок RRU три та чотири логістична регресія не може збільшити P_d, як це робить XGBoost для фіксованої частоти помилкових тривог. Якщо розглянути топологію мережі в розділі 3, де чарунка RRU три має більше сусідів, ніж чарунка RRU чотири, чарунка RRU три є більш мінливою через вищий коефіцієнт передачі обслуговування порівняно з іншими чарунками RRU. Оскільки логістична регресія є менш складною моделлю ML, ніж XGBoost, логістична регресія гірша з прогнозами в обох комірках і не може належним чином покращити показник продуктивності.

Таблиця 4.2 – Середні результати для прогнозів аномалій (%, усереднення виконується за типами пристроїв)

Cell ID	SubsCat	Логістична регресія			XGBoost		
		AUC-ROC	Досто-вірність	Точність	AUC-ROC	Досто-вірність	Точність
1	платина	88,0	55,4	77,8	91,5	63,4	77,5
1	золото	87,4	56,0	77,4	91,5	63,5	77,9
1	срібло	87,6	55,3	77,4	91,7	63,6	78,0
2	платина	87,3	55,5	77,0	91,4	63,3	77,6
2	золото	87,6	55,7	77,5	91,2	63,1	77,6
2	срібло	87,5	56,1	77,5	91,9	63,7	78,0
3	платина	84,9	55,6	75,2	87,7	60,1	75,5
3	золото	85,4	55,8	75,7	88,5	89,8	76,4
3	срібло	84,5	54,9	75,1	87,9	59,4	76,1
4	платина	88,0	56,3	77,5	91,6	63,5	77,7
4	золото	87,4	55,7	77,2	91,4	62,9	77,6
4	срібло	87,9	55,6	76,9	91,8	63,8	77,8
5	платина	87,2	55,5	77,0	91,0	63,1	77,2
5	золото	87,5	55,7	77,2	91,0	63,0	77,3
5	срібло	87,4	55,5	77,1	91,0	63,0	77,6
Середнє		87,0	55,6	76,9	90,7	62,6	77,3

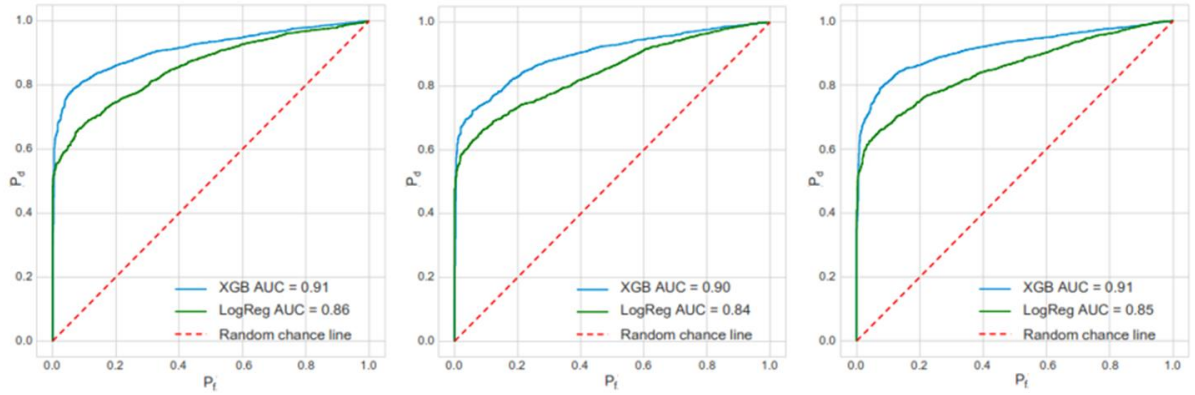
Серед категорій абонентів, як зображено на рисунку 4.2, криві ROC моделей не сильно змінюються. Це через той факт, що категорії передплатників не впливають істотно на продуктивність моделі. З іншого боку, можна побачити, що коли дві криві ROC (тобто криві логістичної регресії та моделі XGBoost) порівнюються в усіх випадках, крива логістичної регресії залишається нижчою за криву XGBoost у кожному окремому випадку, що означає, що XGBoost досягає вищого P_d для даного P_f.



(а) RRU 3, категорія абонента – срібло, тип – мобільний телефон.

(б) RRU 3, категорія абонента – золото, тип – мобільний телефон.

(в) RRU 3, категорія абонента – платина, тип – мобільний телефон.



(г) RRU 4, категорія абонента – срібло, тип – мобільний телефон.

(д) RRU 4, категорія абонента – золото, тип – мобільний телефон.

(е) RRU 4, категорія абонента – платина, тип – мобільний телефон.

Рисунок 4.2 – Частота хибних позитивних результатів порівняно з частотою справжніх позитивних результатів для моделей логістичної регресії та XGBoost (AUC-ROC)

ВИСНОВКИ

Машинне навчання та аналітика даних мають потенціал для революції в мережних операціях, пропонуючи автоматизовані та інтелектуальні рішення для ефективного й точного керування та оптимізації мереж. Використовуючи алгоритми ML, мережні адміністратори можуть завчасно виявляти та вирішувати проблеми з продуктивністю, виявляти збої в мережі та підвищувати безпеку мережі.

Переваги ML і аналітики даних у мережних операціях включають покращену продуктивність мережі, підвищену операційну ефективність, швидше вирішення проблем і покращену безпеку. Однак для успішного впровадження необхідно вирішити такі проблеми, як якість даних, питання конфіденційності, складності інтеграції та потреба в спеціальних навичках.

Реальні приклади демонструють ефективність оптимізації продуктивності мережі та виявлення помилок на основі ML. Ці тематичні дослідження підкреслюють позитивний вплив на надійність мережі, задоволеність клієнтів і ефективність роботи.

Результаті цієї роботи призначені для фахівців у сфері комунікаційних систем і мереж, які планують використовувати аналітичні системи на основі штучного інтелекту для комунікаційних інфраструктур. Основні внески роботи включають таке:

- розглянуто традиційні методики, засновані на навчанні, для NTMA;
- проаналізовано ключові характеристики та застосування NTMA;
- розглянуто методи ML, які використовуються у застосунках NTMA;
- розглянуто NWDAF в архітектурі стільникових мереж 5G;
- запропоновано модель системи для інтелектуальної мережної аналітики у стільникових мережах 5G.

Використано кілька методів ML для вирішення двох основних проблем:

- прогнозування навантаження на мережу за допомогою аналізу часових рядів, зокрема за допомогою моделей лінійної регресії, LSTM і RNN;
- класифікація аномалій в мережі за допомогою моделей логістичної регресії і метода підсилювання градієнта XGBoost.

Крім того, створено набір даних на основі чарунок для аналітики в стільникових мережах 5G з використанням полів, визначених стандартним документом 5G. Із результатів експериментів видно, що моделі нейронної мережі перевершують модель лінійної регресії для правильного прогнозування навантаження на мережу. Подібним чином XGBoost на основі дерева перевершує логістичну регресію під час класифікації аномалій у мережі.

Роботу можна розширити, використовуючи різні моделі AI/ML, зосереджуючись на інших можливостях NWDAF.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Dong, F., Zhang, C., & Cai, W. (2019). An intelligent fault detection and diagnosis method for network operations using machine learning. IEEE Transactions on Network Science and Engineering, 6(4), 508-519.

2. Wang, L., Xu, J., & Zheng, H. (2020). Machine learning-based fault detection and diagnosis in network operations: A survey. Journal of Network and Computer Applications, 166, 102758.

3. Cao, Y., & Wu, Y. (2021). Machine learning-based network fault prediction: A survey. Journal of Network and Computer Applications, 180, 103010.

4. Wang, Y., Huang, X., & Li, Z. (2018). Network anomaly detection based on machine learning techniques: A survey. IEEE Access, 6, 37528-37545.

5. Goyal, R., & Singh, G. (2021). Machine learning in network security: A survey. Computer Communications, 173, 99-119.

6. Islam, R., Hossain, M. S., & Kumar, N. (2021). Machine learning for network slicing in 5G and beyond: Challenges, methodologies, and future directions. IEEE Communications Surveys & Tutorials, 23(1), 305-337.

7. Zhang, Y., et al. (2020). Machine learning and artificial intelligence for network security: A survey. Computer Networks, 173, 107227.

8. Sallahi, F., Hadavi, E., & Al-Fuqaha, A. (2021). Machine learning for edge computing: A comprehensive survey. IEEE Communications Surveys & Tutorials, 23(3), 2355-2401.

9. D Alconzo, A., Drago, I., Morichetta, A., Mellia, M., Casas, P. A survey on big data for network traffic monitoring and analysis. IEEE Trans. Netw. Serv. Manag., 16 (3) (2019), pp. 800-813

10. Shahraki, A., Taherkordi, A., Haugen, Ø., Eliassen, F. Clustering objectives in wireless sensor networks: A survey and research direction analysis. Comput. Netw., 180 (2020), Article 107376

11. Ehrlich, M., Biendarra, A., Trsek, H., Wojtkowiak, E., Jasperneite, J. Passive flow monitoring of hybrid network connections regarding quality of service parameters for the industrial automation. 8 Jahreskolloquium “Kommunikation in der Automation–KommA (2017)

12. Labrinidis, A., Jagadish, H. V. Challenges and opportunities with big data. Proc. VLDB Endow., 5 (12) (2012), pp. 2032-2033

13. Masala, E., Servetti, A., Basso, S., De Martin, J. C. Challenges and issues on collecting and analyzing large volumes of network data measurements. New Trends in Databases and Information Systems, Springer (2014), pp. 203-212

14. Basso, S., Servetti, A., De Martin, J. C. Rationale, design, and implementation of the network neutrality bot. Congresso AICA 2010 (L’Aquila) (2010)

15. Demchenko, Y, De Laat, C., Membrey, P. Defining architecture components of the big data ecosystem. 2014 International Conference on Collaboration Technologies and Systems (CTS), IEEE (2014), pp. 104-112

16. Ricciato, F. Traffic monitoring and analysis for the optimization of a 3g network. IEEE Wirel. Commun., 13 (6) (2006), pp. 42-49

17. Justine, S., Chang, L., Raluca, A. P., Ratnasamy, S. Blindbox: Deep packet inspection over encrypted traffic, in: Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, 2015, pp. 213–226.

18. Marinchak, C., Forrest, E., Hoanca, B. The impact of artificial intelligence and virtual personal assistants on marketing. Encyclopedia of Information Science and Technology, Fourth Edition, IGI global (2018), pp. 5748-5756

19. Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. MIT press (2016)

20. Konečnỳ, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A., Bacon D. Federated learning: Strategies for improving communication efficiency. (2016). arXiv preprint arXiv:1610.05492

21. “Mobile visual networking index (VNI),” accessed: 2019-12-05. [Online]. Available: https://www.cisco.com/c/m/en_us/solutions/service-provider/visual- networking- index.html

22. Andrews, J. G., Buzzi, S., Choi, W., Hanly, S. V., Lozano, A., Soong, A. C., Zhang, J. C. “What will 5G be?” IEEE J. Selected Areas Commun., vol. 32, no. 6, pp. 1065-1082, June 2014.

23. “Non-standalone and standalone: Two standards-based paths to 5G,” accessed: 2019-12-05. [Online]. Available: https://www.ericsson.com/en/blog/2019/7/standalone- and-non-standalone-5g-nr-two-5g-tracks

24. 3GPP, “5G System; Unified data management services; Stage 3,” 3rd Generation Partnership Project (3GPP), Technical Specification (TS 29.503), Sept. 2019, version 16.1.0. [On line]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3342

25. Goodfellow, I. Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. “Generative adversarial nets,” in Proc. NIPS, 2014, pp. 2672-2680.

26. Sevgican, S., Turan, M., Gokarslan, K., Yilmaz, H. B., T. Tugcu, “Synthetic 5G cellular network data for NWDAF,” 2019. [Online]. Available: https://github.com/sevgicansalih/nwdaf_data.

27. Лушпа Б.Є., Куриленко А.О., Янковський О.А. Управління трафіком мереж // Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління : тези доп. 13-ї міжнар. наук.-техн. конф., Баку - Харків - Жиліна : [у 2 т.]. Т.2: секція 2 / Нац. ун-т оборони Азербайджанської Республіки [та ін.]. – Харків : Impress, – 2023. – С. 103.