

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА

Навчально-науковий інститут
«Інститут інформаційних технологій в економіці»

Кафедра математичного моделювання та статистики

Освітньо-професійна програма «Економічна кібернетика і Дата Сайнс»

Галузь знань 05 «Соціальні та поведінкові науки»
Спеціальність 051 «Економіка»

Форма навчання: очна (денна)

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА

на тему «Прогнозування схильності до інтернет залежності методами Data Science»

(назва теми)

здобувача Василенко Денис Павловича

(ПІБ, підпис)

Науковий керівник: кандидат економічних наук, доцент, Ярослава ІЩУК

(підпис)

Робота допущена до захисту перед екзаменаційною
комісією з атестації здобувачів вищої освіти (ЕК)

Завідувач кафедри: к. ф.-м. н., професор
Галина ВЕЛИКОІВАНЕНКО

(підпис)

Київ 2024

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА**

**Навчально-науковий інститут
«Інститут інформаційних технологій в економіці»**

Кафедра математичного моделювання та статистики

Освітньо-професійна програма «Економічна кібернетика і Дата Сайнс»

Галузь знань 05 «Соціальні та поведінкові науки»
Спеціальність 051 «Економіка»

ПОГОДЖЕНО

Керівник проектної групи (гарант)
освітньо-професійної програми
Олена ПІСКУНОВА

(підпис)

2024 р.

ЗАТВЕРДЖУЮ

Завідувач кафедри математичного
моделювання та статистики
Галина ВЕЛИКОІВАНЕНКО

(підпис)

2024 р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

здобувача вищої освіти

Василенка Дениса Павловича
(прізвище, ім'я, по батькові)

очної (денної)

очної (денної), заочної, дистанційної

форми навчання

на підготовку кваліфікаційної магістерської роботи

на тему

«Прогнозування схильності до інтернет залежності методами Data Science»

Тему затверджено наказом ректора Університету від «19» вересня 2024 р. № 1610 / ст

**Кваліфікаційна магістерська робота виконується на матеріалах
платформи Kaggle**

План кваліфікаційної магістерської роботи

Розділ 1	Аналіз інтернет залежності як соціальної проблеми
Розділ 2	Застосування методів Data Science для аналізу інтернет залежностей
Розділ 3	Прогнозування інтернет залежності методами машинного навчання

Об'єкт дослідження:	процес формування інтернет залежності у дітей і підлітків
Предмет дослідження:	методи Data Science для аналізу і прогнозування інтернет залежності.
Мета виконання кваліфікаційної магістерської роботи:	дослідження проблеми інтернет залежності як актуального поведінкового розладу, ідентифікація ключових факторів, які сприяють розвитку цієї залежності, та розробка методу прогнозування схильності індивіда до інтернет залежності в умовах відсутності експертизи

Конкретні завдання, які здобувач повинен виконати для досягнення поставленої мети:

У розділі 1: Визначити поняття інтернет залежності та дослідити основні підходи до її трактування у сучасній науковій літературі; дослідити основні соціальні аспекти інтернет-залежності та виявити її вплив на різні групи населення; виявити фактори, що сприяють розвитку інтернет залежності, включаючи індивідуальні, соціальні, культурні та технологічні передумови.

У розділі 2: Визначити методи очищення, нормалізації та трансформації даних, зібраних для аналізу інтернет залежностей; розробити алгоритм підготовки даних для моделювання; обґрунтувати вибір алгоритмів машинного навчання; провести порівняння продуктивності обраних моделей; визначити особливості роботи моделей

У розділі 3 Обґрунтувати вибір методів машинного навчання для задач прогнозування інтернет-залежності, враховуючи специфіку доступних даних і дослідницьку мету; провести підготовку та попередню обробку даних; побудувати та навчити базові моделі машинного навчання для прогнозування ризику інтернет-залежності; провести порівняльний аналіз продуктивності моделей; застосувати крос-валідацію для оцінки узагальнюючої здатності моделей

Завдання підготував
науковий керівник

(підпис)

Ярослава ІЩУК

(ім'я, прізвище)

«__» _____ 2024 р.

Завдання одержав
здобувач

(підпис)

Денис ВАСИЛЕНКО

(ініціали, прізвище)

«__» _____ 2024 р.

Реферат

Кваліфікаційна магістерська робота містить 73 сторінки, 10 рисунків, перелік використаних джерел з 60 найменувань.

«Прогнозування схильності до інтернет залежності методами Data Science»

Об'єктом дослідження є процес формування інтернет залежності.

Предметом дослідження є методи Data Science для аналізу і прогнозування інтернет залежності.

Метою роботи є дослідження проблеми інтернет залежності як актуального поведінкового розладу, ідентифікація ключових факторів, які сприяють розвитку, та розробка методу прогнозування схильності індивіда до інтернет залежності з метою виявлення проблеми в умовах з відсутністю експертизи.

Відповідно до поставленої мети визначені такі завдання:

- дослідити явища інтернет залежності на суб'єктивному і соціальному рівні;
- визначити фактори які впливають на появу розлада;
- оглянути методи боротьби з залежністю;
- проаналізувати методи обробки даних і покращення результатів прогнозування;
- сформувати прогнозну модель машинного навчання на основі градієнтного бустингу;
- розробити алгоритм прогнозування схильності до інтернет залежності і його оцінка;

Теоретична, методична та практична значущість отриманих результатів. Теоретична значущість дослідження полягає у визначенні та аналізі ключових факторів, які впливають на розвиток інтернет залежності, а також у застосуванні алгоритмів градієнтного бустингу, автоенкодерів і моделей для обробки часових рядів.

Методична значущість роботи проявляється у запропонованій структурі обробки та аналізу даних, включаючи процедури передобробки, заповнення пропусків, побудови нових ознак та інтеграції даних акселерометрів.

Практична значущість роботи полягає у створенні прогнозної моделі, яка може бути використана для ранньої ідентифікації індивідів із підвищеним ризиком інтернет залежності. Результати дослідження можуть бути застосовані як у конкретних освітніх або медичних установах для формування інтервенційних програм, так і в ширшому масштабі для розробки політик із профілактики залежності на національному або глобальному рівнях.

Рік виконання кваліфікаційної магістерської роботи – 2024.

Рік захисту роботи – 2024.

Ключові слова: інтернет-залежність, прогнозування, машинне навчання, Data Science, градієнтний бустинг, LightGBM, XGBoost, CatBoost автоенкодер, часові ряди, feature engineering, соціальні фактори, поведінкові розлади, когнітивно-поведінкова терапія.

В і д г у к
про кваліфікаційну магістерську роботу
здобувача освітньо-професійної програми «Економічна кібернетика і Дата
Сайнс»
навчально-наукового інституту
«Інститут інформаційних технологій в економіці»
Василенка Дениса Павловича
на тему «**Прогнозування схильності до інтернет залежності методами Data
Science**»

Актуальність теми. Зростання поширеності Інтернету та його вплив на життя суспільства роблять актуальним розробку ефективних методів прогнозування інтернет-залежності для своєчасної профілактики та лікування. Використання методів Data Science для аналізу великих обсягів даних дозволяє створювати точні моделі прогнозування схильності до інтернет-залежності, що є важливим для розробки персоналізованих програм реабілітації.

Позитивні риси кваліфікаційної магістерської роботи: Магістерська робота є комплексним дослідженням проблеми інтернет-залежності, яке поєднує в собі глибокий теоретичний аналіз та практичне застосування методів Data Science. Результати дослідження можуть бути використані для розробки ефективних стратегій профілактики та лікування цієї актуальної соціальної проблеми. Крім того робота демонструє успішну інтеграцію знань з психології, соціології та економічної кібернетики. Завдяки такому міждисциплінарному підходу вдалося провести всебічний аналіз проблеми інтернет-залежності та розробити ефективні інструменти для її вирішення.

Наявність самостійних розробок автора Робота відзначається високим рівнем наукової новизни завдяки застосуванню сучасних методів Data Science для прогнозування інтернет-залежності. Детальний аналіз даних та розробка ефективних моделей машинного навчання дозволили отримати цінні результати, які можуть бути використані для подальших досліджень у цій галузі

Цінність теоретичних висновків та практичних рекомендацій. Результати дослідження мають високий потенціал для практичного застосування. Розроблені моделі прогнозування можуть бути використані для виявлення інтернет залежності методами Data Science без допомоги фахівців-психологів, створення індивідуальних програм реабілітації, а також для розробки державної політики у сфері боротьби з інтернет-залежністю. Результати дослідження пройшли апробацію на 91 щорічній студентській науковій конференції «Інноваційні проекти для економічного відродження та конкурентного розвитку України» та на XII всеукраїнській науково-практичній конференції форумі молодих економістів-кібернетиків «Моделювання економіки: проблеми, тенденції, досвід»

Наявність недоліків: Додавання кроку кластеризації до дослідження могло б значно підвищити його якість та дозволило б отримати більш деталізовані та точні результати, врахувати специфічні особливості кожної групи. Є окремі недоліки та некоректності в опрацюванні отриманих результатів. До вже існуючих недоліків можна додати відсутність достатньої візуалізації результатів, що обмежує цінність

дослідження.

Загальна оцінка кваліфікаційної магістерської роботи та її допущення до захисту перед ЕК: Кваліфікаційна магістерська робота виконана на позитивному рівні та свідчить про достатній рівень сформованості компетентностей її автора, передбачених освітньою-професійною програмою «Економічна кібернетика і Дата Сайнс». Вважаю, що робота може бути допущена до захисту перед ЕК.

Науковий керівник

доцент кафедри математичного моделювання та статистики, к.е.н. Ярослава ІЩУК

Науковий керівник доцент кафедри математичного моделювання та статистики, к.е.н

_____ Ярослава ІЩУК _____
(ім'я, ПРІЗВИЩЕ)

_____ (підпис)
“ ____ ” _____ 20__ р.

Рецензія

на кваліфікаційну магістерську роботу
здобувача вищої освіти

Василенка Дениса Павловича

Тема «Прогнозування схильності до інтернет залежності методами Data Science»

Робота, присвячена прогнозуванню схильності до інтернет-залежності методами Data Science, є важливим внеском у дослідження сучасних викликів, пов'язаних із надмірним використанням інтернету серед дітей і підлітків. У роботі розглянуто комплексний підхід до аналізу багатовимірних даних, які включають демографічні, поведінкові, фізичні та психологічні показники. Автор використовує передові методи машинного навчання, зокрема градієнтний бустинг (LightGBM, XGBoost, CatBoost), що забезпечують високу точність моделювання навіть на складних і неоднорідних вибірках.

Значущість роботи обґрунтована актуальністю проблеми, оскільки інтернет-залежність є поширеним явищем, яке негативно впливає на фізичне, психічне та соціальне здоров'я. Зокрема, використання фізичних показників, таких як індекс маси тіла, рівень активності та показники здоров'я, у поєднанні з поведінковими змінними (кількість годин за комп'ютером) дозволяє зробити більш точні прогнози та сформувані науково обґрунтовані рекомендації для профілактики залежності.

Серед основних переваг роботи варто виділити системний підхід до обробки даних. Автор демонструє глибоке розуміння етапів препроцесінгу, таких як обробка пропущених значень, масштабування та імітація, що підвищує якість моделей. Більш того, використання ансамблевих методів градієнтного бустингу забезпечило максимальну ефективність у моделюванні залежності між численними змінними та цільовим показником (Severity Impairment Index). Також варто відзначити візуалізацію результатів, яка значно полегшує інтерпретацію отриманих даних і дозволяє виявляти ключові закономірності.

Робота демонструє потенціал для практичного застосування. Модель, побудована в дослідженні, може бути використана для раннього виявлення груп ризику та створення превентивних заходів, спрямованих на зниження рівня інтернет-залежності серед молоді. Використання доступних і широко поширених змінних, таких як фізична активність і поведінкові патерни, робить запропоноване рішення масштабованим і придатним для різних контекстів.

Проте в роботі можна відзначити певний недолік. Хоча автор детально аналізує кореляції та будує моделі, меншою мірою увага приділена інтерпретації отриманих результатів у соціальному контексті. Зокрема, в дослідженні не було детально розглянуто, як показники інтернет залежності можуть варіюватися в різних соціальних групах або регіонах. Цей аспект, хоча й не є центральним для математичного моделювання, міг би підсилити висновки роботи в прикладному сенсі.

Загалом, робота демонструє високий рівень методологічної підготовки та

актуальність результатів. Запропоновані підходи мають потенціал для практичного впровадження в сфері охорони здоров'я та освітніх ініціатив, спрямованих на зменшення шкоди від надмірного використання інтернету.

Горна Марина Олексіївна

доцент кафедри статистики,

інформаційно-аналітичних систем і демографії,

Київський національний університет імені Тараса Шевченка

ЗМІСТ

ВСТУП.....	3
РОЗДІЛ 1 АНАЛІЗ ІНТЕРНЕТ ЗАЛЕЖНОСТІ ЯК СОЦІАЛЬНОЇ ПРОБЛЕМИ ...	6
1.1 Методи діагностики та основні прояви інтернет залежності	6
1.3. Передумови виникнення залежності серед різних вікових груп	16
1.4. Методи боротьби з інтернет залежністю: міжнародний досвід	21
РОЗДІЛ 2 ЗАСТОСУВАННЯ МЕТОДІВ DATA SCIENCE ДЛЯ АНАЛІЗУ ІНТЕРНЕТ ЗАЛЕЖНОСТЕЙ	27
2.1. Методика попередньої обробки даних для прогнозування	27
2.2. Методика прогнозування інтернет залежності на основі методів машинного начання	39
2.3. Методи вдосконалення навчання моделей і оцінки їхньої ефективності	47
РОЗДІЛ 3 ПРОГНОЗУВАННЯ ІНТЕРНЕТ ЗАЛЕЖНОСТІ МЕТОДАМИ МАШИННОГО НАВЧАННЯ.....	50
3.1. Характеристика даних: структура та джерела	50
3.2. Попередня обробка даних: методи очищення та аналіз	58
3.3. Тренування моделей: алгоритми, параметри та оцінка ефективності.....	63
ВИСНОВКИ	70
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:	74

ВСТУП

Інтернет залежність є сучасним поведінковим розладом, що характеризується надмірним та неконтрольованим використанням інтернету, яке призводить до значного порушення соціального, професійного чи особистого функціонування. Згідно з теоріями поведінкової психології, це явище можна класифікувати як одну з форм ненаркотичної залежності, де об'єктом адиктивної поведінки є не речовина, а специфічний вид діяльності, зокрема активність у віртуальному просторі.

Актуальність дослідження інтернет залежності зумовлена її зростаючим впливом на психічне здоров'я, соціальну адаптацію та продуктивність людей, особливо дітей і підлітків, які є найбільш вразливими до цього розладу. У сучасному цифровому світі, де інтернет є невід'ємною частиною повсякденного життя, недостатність діагностичних інструментів та профілактичних заходів ускладнює раннє виявлення та подолання цієї проблеми. Це створює потребу в розробці ефективних методів прогнозування, які базуються на об'єктивних даних, для вчасного втручання і корекції залежності.

В аналізі останніх публікацій і досліджень з прогнозування інтернет залежності активно використовуються методи машинного навчання для аналізу факторів ризику та моделювання складних взаємозв'язків. Чжицзун Го та Лайган Чжан у своїй роботі представили модель на основі Deep Belief Network (DBN), яка дозволяє точно прогнозувати схильність до інтернет залежності через багатофакторний аналіз, показуючи високу точність і ефективність. Хуа Юмін і Го Цзюньхай у своїй публікації розглянули загальні принципи глибинного навчання та DBN, що заклало теоретичну основу для подальших прикладних досліджень у цій галузі. Шуанг Чжан і Хуісі Ю запропонували gd-LSTM модель для аналізу поведінкових даних студентів, яка показала переваги у точності прогнозів інтернет залежності. Крім того, робота Чжана та інших показала ефективність gd-LSTM у контексті прогнозування часових рядів, що підкреслює потенціал цієї моделі для

складних аналітичних завдань. Усі ці роботи створюють наукову основу для подальшого вдосконалення моделей, зокрема їх адаптації до різноманітних факторів та умов.

Метою роботи є дослідження проблеми інтернет залежності як актуального поведінкового розладу, ідентифікація ключових факторів, які сприяють розвитку, та розробка методу прогнозування схильності індивіда до інтернет залежності з метою виявлення проблеми в умовах з відсутністю експертизи.

Завданням роботи є реалізація мети, що полягає в виконанні наступних завдань:

- дослідити явища інтернет залежності на суб'єктивному і соціальному рівні;
- визначити фактори які впливають на появу розлада;
- оглянути методи боротьби с залежністю;
- проаналізувати методи обробки даних і покращення результатів прогнозування;
- сформувати прогнозну модель машинного навчання на основі градієнтного бустингу;
- розробити алгоритм прогнозування схильності до інтернет залежності і його оцінка;

Об'єктом дослідження є процес формування інтернет залежності.

Предметом дослідження є методи Data Science для аналізу і прогнозування інтернет залежності.

Методами дослідження виступають прогнозні моделі на основі методів градієнтного бустингу LightGBM, XGBoost і CatBoost, результати прогнозування яких були об'єднані шляхом Voting Regression, і методи попередньої обробки з метою покращення результатів, такі як автоенкодер, time series aggregation, feature engineering і dimensionality reduction.

Значущість отриманих результатів цієї кваліфікаційної роботи полягає в розробці науково обґрунтованого підходу до прогнозування схильності до інтернет залежності з використанням сучасних методів машинного навчання.

Теоретична значущість дослідження полягає у визначенні та аналізі ключових факторів, які впливають на розвиток інтернет залежності, а також у застосуванні алгоритмів градієнтного бустингу, автоенкодерів і моделей для обробки часових рядів.

Методична значущість роботи проявляється у запропонованій структурі обробки та аналізу даних, включаючи процедури передобробки, заповнення пропусків, побудови нових ознак та інтеграції даних акселерометрів.

Практична значущість роботи полягає у створенні прогнозної моделі, яка може бути використана для ранньої ідентифікації індивідів із підвищеним ризиком інтернет залежності. Результати дослідження можуть бути застосовані як у конкретних освітніх або медичних установах для формування інтервенційних програм, так і в ширшому масштабі для розробки політик із профілактики залежності на національному або глобальному рівнях.

Інформаційна база дослідження охоплює праці зарубіжних науковців, зокрема роботи, присвячені теоретичному аналізу інтернет залежності, сучасним методам машинного навчання та їх застосуванню. Використано статистичні дані, офіційні звіти міжнародних організацій і результати практичних досліджень, які забезпечили основу для комплексного аналізу та розробки прогностичних моделей.

Структура роботи. Робота складається зі вступу, трьох розділів, висновків, переліку використаних джерел.

РОЗДІЛ 1

АНАЛІЗ ІНТЕРНЕТ ЗАЛЕЖНОСТІ ЯК СОЦІАЛЬНОЇ ПРОБЛЕМИ

1.1 Методи діагностики та основні прояви інтернет залежності

Зародження інтернету та його поширення у глобальному масштабі стало однією з ключових віх у розвитку сучасної цивілізації, трансформувавши не лише спосіб обміну інформацією, а й саму структуру суспільства. Початок цього технологічного прориву було покладено у 1960-х роках, коли в рамках проєкту ARPANET, фінансованим Агентством передових дослідницьких проєктів Міністерства оборони США, було створено першу комп'ютерну мережу, здатну забезпечувати обмін даними між географічно розташованими вузлами. [6] Ця система мала на меті підтримання комунікацій у випадку катастроф або воєнних дій, що продиктовано контекстом Холодної війни.

Інженери та дослідники, які працювали над ARPANET, зосередили свої зусилля на розробці пакетної комутації, що стало фундаментальною технологією для ефективної передачі інформації. У 1970-х роках цей підхід розвинувся у створення протоколу TCP/IP, який заклав основи для глобальної взаємодії комп'ютерних мереж. У 1980-х роках розробка та впровадження електронної пошти, віддаленого доступу та інших технологій поступово залучали до мережі наукові установи та університети по всьому світу.

Перехід від обмеженої мережі для наукових і військових цілей до всесвітньо доступної системи став можливим завдяки створенню у 1989 році Всесвітнього павутиння (World Wide Web) британським ученим Тімом Бернерсом-Лі. Його ідея про використання гіпертексту як засобу для організації інформації дозволила зробити інтернет інтуїтивно зрозумілим та доступним для широкої аудиторії. У 1990-х роках із комерціалізацією інтернету та розвитком браузерів і провайдерів доступу розпочався новий етап його еволюції, позначений стрімким зростанням кількості користувачів.

Значущість цього технологічного прориву важко переоцінити, адже він став каталізатором численних змін у суспільстві. Інтернет не лише спростив доступ до знань, але й створив нові форми економічної активності, культурної взаємодії та соціального обміну. Водночас він започаткував еру глибокої цифровізації, що сформувала новий стиль життя, де інтеграція технологій у повсякденність стала нормою. Це заклало передумови для позитивних і негативних наслідків, зокрема появи інтернет залежності як нового соціально-психологічного виклику.

Поява інтернету як соціального феномену стала визначальним моментом у формуванні нових моделей комунікації, освіти та доступу до інформації, що докорінно змінили соціальну структуру і спосіб життя сучасної людини. Якщо на початку свого розвитку інтернет виконував переважно функцію наукової та військової інфраструктури, то в середині 1990-х років його комерціалізація та швидке поширення в побутовій сфері спричинили глобальні зміни у взаємодії між людьми, трансформуючи її у новий формат, позбавлений географічних обмежень.

Інтернет відкрив можливість для миттєвого обміну інформацією через електронну пошту, чати та форуми, які стали першим кроком у переході до цифрового спілкування. Цей процес кардинально змінив динаміку людської взаємодії, зменшивши значущість фізичної присутності та створивши новий рівень соціальної доступності. Одночасно із цим зростання доступу до всесвітньої павутини стимулювало розвиток цифрових освітніх платформ і ресурсів, які зробили знання більш демократичними та доступними для широкого загалу. Від онлайн-бібліотек до дистанційних навчальних програм — інтернет став інструментом, що забезпечив рівний доступ до інформації для людей з будьяким соціальним статусом чи географічним розташуванням.

Ключовим етапом у становленні інтернету як соціального явища стала поява інтерактивних сервісів наприкінці 1990-х — початку 2000-х років. Соціальні мережі, такі як Friendster, MySpace і пізніше Facebook, створили безпрецедентну платформу для соціалізації, де користувачі могли обмінюватися думками, фотографіями та подіями свого життя в реальному часі. Одночасно розвивалися

онлайн-ігри, які об'єднали мільйони гравців у віртуальних світах, формуючи новий тип взаємодії, що базується на спільній участі та емоційному залученні.

Також важливою складовою цього явища стало стрімке зростання популярності блогів і відеохостингів, таких як YouTube, які створили новий формат самовираження і культурного обміну. Всі ці сервіси заклали основу для сучасного цифрового світу, що характеризується глибокою інтерактивністю, персоналізацією контенту та інтеграцією технологій у всі сфери життя.

Інтернет як соціальний феномен став не лише джерелом можливостей, але й викликів, створивши умови для надмірного використання технологій і формування залежності. Його вплив на комунікацію, освіту та доступність інформації не лише відкрив нові горизонти для людства, але й поставив нові питання перед суспільством щодо балансування між цифровими та реальними аспектами життя.

Зі стрімким поширенням інтернету в 1990-х роках суспільство почало стикатися з феноменом, коли надмірне використання всесвітньої павутини перетворювалося на патологічну залежність, що мала серйозні психологічні, соціальні та фізичні наслідки. Це явище швидко привернуло увагу дослідників, які намагалися осмислити його причини, природу та можливі підходи до діагностики й терапії.

Термін «інтернет залежність» вперше увійшов до наукового обігу завдяки роботі американської психологині Кімберлі Янг, яка у 1996 році представила одну з перших систематичних спроб дослідити це явище. Її публікація "Internet Addiction: The Emergence of a New Clinical Disorder" у журналі *CyberPsychology & Behavior* не лише ввела термін у науковий дискурс, але й запропонувала концептуальні підходи до розуміння цієї форми поведінкової залежності [7]. У своїх дослідженнях Янг адаптувала критерії діагностики патологічної азартної гри, визначені в DSM-IV, для аналізу інтернет залежності, що стало основою для створення подальших інструментів діагностики [8].

Янг звернула увагу на такі аспекти, як неконтрольована потреба у використанні інтернету, емоційний дискомфорт у випадку відсутності доступу до мережі, а також значні соціальні, професійні чи освітні проблеми, які виникають

внаслідок надмірного часу, проведеного онлайн. Її дослідження ґрунтувалося на емпіричних даних, отриманих через анкетування, що дозволило створити перші профілі користувачів, схильних до інтернет залежності.

Важливість роботи Янг полягає не лише у визначенні основних характеристик інтернет залежності, але й у започаткуванні наукового діалогу щодо цього феномену. Її дослідження викликали інтерес серед психологів, соціологів і нейропсихологів, сприяючи розвитку міждисциплінарних підходів до вивчення проблеми. Завдяки її ініціативам інтернет залежність почала розглядатися як окрема клінічна проблема, що потребує уваги з боку медицини, психології та соціальних наук.

Таким чином, праці Кімберлі Янг стали відправною точкою для формування нового напрямку в наукових дослідженнях, заклавши підґрунтя для подальшого розуміння, класифікації та розробки методів подолання інтернет залежності. Важливість цих досліджень зростала разом із поширенням інтернету, актуалізуючи проблему та забезпечуючи наукове обґрунтування для її подальшого вивчення.

Починаючи з 2000-х років, поширення мобільного інтернету та доступність смартфонів докорінно змінили характер використання цифрових технологій, зробивши інтернет постійно доступним. Ця трансформація суттєво посилила ризики розвитку залежності, оскільки миттєвий доступ до онлайн-ресурсів та інтерактивних платформ стимулював формування нових патернів поведінки, які характеризуються високою частотою та інтенсивністю використання. Мобільні пристрої, оснащені все більш потужними апаратними і програмними засобами, зруйнували бар'єр просторової прив'язаності, властивий стаціонарному використанню комп'ютерів. Завдяки цьому інтернет проник у всі сфери повсякденного життя, від робочого середовища до дозвілля. Зокрема, розвиток додатків для миттєвих повідомлень, соціальних мереж, онлайн-ігор і стрімінгових платформ стимулював формування гіперзв'язку — стану постійної включеності у цифрове середовище. Водночас, цей феномен сприяв виникненню поведінкових патернів, які базуються на потребі у постійному інформаційному стимулюванні, що є ключовим тригером інтернет залежності.

Особливо критичними для еволюції залежності стали 2010-ті роки, коли технологічний прогрес забезпечив безперервне підключення до інтернету за рахунок розвитку 4G-мереж, а згодом і появи 5G. Смартфони, оснащені високоякісними сенсорними екранами та функціональністю, яка конкурує з комп'ютерами, стали універсальними інструментами для роботи, навчання та розваг. Однак їхня всепроникність також призвела до зміщення акценту від епізодичного до постійного використання. Звичка перевіряти смартфон кілька разів на годину, постійно оновлювати стрічку соціальних мереж або грати в мобільні ігри сформувала новий тип поведінкової залежності, яка часто проходить непомітно для користувачів.

Суттєвим фактором посилення інтернет залежності стала розробка алгоритмічних рішень для персоналізації контенту. Соціальні мережі та онлайн-додатки використовують потужні інструменти аналізу даних, що дозволяє формувати індивідуальні стрічки новин, які максимізують час перебування користувачів онлайн. Це викликає унікальні психологічні механізми, що базуються на негайному задоволенні, яке підкріплює компульсивне використання інтернету. Зміщення від комп'ютерного до мобільного формату споживання контенту також стимулює виникнення нових соціокультурних норм, де безперервна онлайн-присутність сприймається як необхідна умова для інтеграції у сучасне суспільство [9].

Еволюція інтернет залежності є результатом взаємодії технологічного прогресу, соціальних змін та людської психології. Глобалізація мобільного інтернету та гіперзв'язок створили безпрецедентні можливості для доступу до інформації, але водночас значно підвищили ризики формування поведінкових залежностей, які набувають все більшого поширення у сучасному суспільстві.

Перші прояви інтернет залежності як соціального феномену були тісно пов'язані з розвитком культур віртуальної взаємодії, які сформувалися внаслідок появи інтернет кафе, онлайн-ігор та ранніх комунікаційних платформ, таких як чати і форуми. Наприкінці 1990-х — на початку 2000-х років інтернет кафе стали важливими осередками суспільної взаємодії, зокрема серед молоді та підлітків,

пропонуючи доступ до глобальної мережі, який у той час ще не був доступний у багатьох домогосподарствах. Ці заклади стали своєрідними хабами для нової цифрової культури, яка поєднувала розваги, комунікацію та навчання.

Особливої популярності набула практика колективного використання інтернету через участь в онлайн-іграх, що дозволяло не лише отримати індивідуальне задоволення, але й створювало умови для формування соціальних зв'язків у віртуальному середовищі. Однак ці зв'язки часто базувалися на конкурентних і емоційно інтенсивних взаємодіях, які стимулювали тривале перебування у мережі. Ігрові платформи вперше виявили здатність цифрового середовища формувати поведінкові звички, що характеризувалися високою частотою та нав'язливістю.

В той самий час інтернет чати та форуми стали одними з перших платформ для текстової комунікації в реальному часі, які надавали можливість створювати нові форми соціальних відносин, що не залежали від географічних меж. Ці технології, хоч і сприяли демократизації доступу до інформації та відкривали нові горизонти для самовираження, також створювали ризики для соціальної ізоляції. Учасники таких спільнот часто замінювали реальні міжособистісні контакти на віртуальні, що призводило до зміщення балансу між офлайн і онлайн взаємодіями [10].

Значний вплив на соціальну поведінку мала і специфіка ранніх інтернет ресурсів, які були орієнтовані на створення анонімних або псевдоанонімних профілів. Це сприяло зниженню соціальних бар'єрів та створювало умови для експериментування з власною ідентичністю, але також стимулювало надмірне залучення, оскільки користувачі прагнули постійно підтримувати свою присутність у мережі. Така динаміка особливо вплинула на молодь, яка є найбільш сприйнятливою до формування звичок у період інтенсивного соціального та емоційного розвитку.

Ранні прояви інтернет залежності також характеризувалися помітною сегментацією за соціальними групами. Наприклад, підлітки та студенти, які активно використовували інтернет для розваг та навчання, демонстрували високий

ризик формування залежності, тоді як старші покоління, які залучалися до мережі для роботи чи комунікації, значно рідше виявляли подібні тенденції [11]. У цьому контексті інтернет кафе стали символом культурної зміни, що розпочала перетворення сприйняття технологій як інструмента збагачення соціального життя, але водночас посилила ризики психологічної та поведінкової залежності. Поява віртуальних платформ створила нові форми соціального життя, але також заклала основу для ризиків, які сьогодні розглядаються як важливі виклики для психології, соціології та цифрової культури.

Інтернет залежність у XXI столітті перетворилася на глобальну проблему, що охоплює всі соціальні верстви та регіони світу. Цей феномен виник на тлі масової доступності цифрових технологій, зокрема мобільного інтернету, що забезпечив можливість постійного зв'язку та інтерактивної діяльності. Однак зростання часу, проведеного в мережі, стало чинником ризику для фізичного, психічного та соціального здоров'я мільйонів людей, створюючи умови для формування різних форм залежності, таких як патологічний геймінг, компульсивне користування соціальними мережами чи надмірне споживання інформації.

Цей тип залежності вийшов за рамки локальних явищ і став предметом міжнародного дискурсу, що визнається науковою спільнотою та урядами як значний виклик сучасності. Заходи, спрямовані на боротьбу з інтернет залежністю, ініційовані Всесвітньою організацією охорони здоров'я (ВООЗ) та іншими міжнародними органами, включають спроби стандартизувати діагностику, визначити критерії оцінювання впливу та впровадити профілактичні заходи [12].

Проблема інтернет залежності поглиблюється глобалізацією економіки та інтеграцією цифрових технологій у всі аспекти життя. Економічна конкуренція змушує компанії створювати дедалі привабливіші платформи для онлайн-взаємодії, від соціальних мереж до платформ для геймінгу, спрямованих на утримання уваги користувачів. Соціальна ізоляція під час пандемії COVID-19 стала катализатором переходу суспільства до переважно цифрового способу взаємодії, зробивши інтернет основним інструментом для роботи, навчання й комунікації.

Ця трансформація призвела до нового рівня психологічного та соціального стресу, посилюючи схильність до залежності через постійний доступ до інтерактивних платформ. Залежність формується через механізми позитивного підкріплення, властиві багатьом сучасним сервісам, зокрема алгоритмам, які адаптують контент до уподобань користувача.

Таким чином, інтернет залежність стала явищем, яке має вплив не лише на індивідуальному, а й на глобальному рівні, вимагаючи координації зусиль між країнами та адаптації стратегій до специфіки кожного регіону. Ефективне вирішення цієї проблеми потребує інтегрованих моделей, які враховують як культурні, так і технологічні аспекти сучасного суспільства.

Критерії діагностики інтернет залежності базуються на міжнародних класифікаціях психічних розладів, таких як DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, 5-те видання) та ICD-11 (Міжнародна класифікація хвороб, 11-те видання). DSM-5 не визнає інтернет залежність як окремий розлад, але включає "Розлад інтернет ігор" (Internet Gaming Disorder) у розділі III для подальшого вивчення, визначаючи дев'ять критеріїв, серед яких: постійна зайнятість іграми, симптоми відміни, невдалі спроби контролю тощо. Діагноз встановлюється за наявності п'яти критеріїв протягом 12 місяців. ICD-11 офіційно визнає "Геймінговий розлад" як поведінкову залежність із трьома ключовими симптомами: порушенням контролю, пріоритетом ігор над іншими активностями та продовженням поведінки попри негативні наслідки [13][14].

Інтернет залежність характеризується втратою контролю над часом онлайн, емоційною залежністю (ейфорія, дратівливість без доступу), уникненням офлайн-обов'язків, що призводить до соціальної ізоляції та зниження продуктивності. Нейробіологічні дослідження виявляють зміни у мозкових структурах, пов'язаних із винагородою та імпульсами, подібні до змін при залежності від речовин [8].

Для діагностики використовуються методи тестування, зокрема Internet Addiction Test (IAT), що оцінює залученість і контроль, а також аналіз біометричних та поведінкових даних (електроенцефалографія, патерни

використання інтернету). Методи машинного навчання дозволяють прогнозувати ризик розвитку залежності [15].

Інтернет залежність має різноманітні форми: залежність від соціальних мереж (постійне перевіряння оновлень, зниження самооцінки), онлайн-ігор (надмірне занурення у MMORPG), відеоплатформ (перегляд відео за рахунок продуктивності) та онлайн-шопінгу (компульсивні покупки). Культурні та вікові фактори впливають на прояви залежності: підлітки частіше схильні до соціальних мереж і онлайн-ігор, дорослі — до шопінгу й відеоплатформ. У країнах із високим технологічним розвитком інтернет залежність поширеніша, а культурні цінності визначають, які онлайн-активності є привабливими.

Онлайн-шопінгова залежність, або компульсивний розлад покупок, характеризується неконтрольованим бажанням здійснювати покупки через інтернет, що спричиняє фінансові труднощі та психологічний стрес. Часто така поведінка пов'язана з тривожністю чи депресією, де шопінг виступає механізмом подолання. Ця форма залежності поширена серед дорослих, а її прояви зумовлені культурними факторами, зокрема споживацтвом та ставленням до матеріальних благ.

Вікові відмінності проявів інтернет залежності очевидні: молодші особи більше схильні до залежності від соціальних мереж та ігор, тоді як дорослі частіше демонструють залежність від онлайн-шопінгу чи азартних ігор. Це зумовлено віковими особливостями, такими як пошук ідентичності у підлітковому віці або подолання стресу у дорослих [16].

Культурні фактори також суттєво впливають на формування інтернет залежності. У суспільствах, де онлайн-ігри соціально прийнятні, вища поширеність геймінгової залежності, тоді як у культурах із суворими обмеженнями на використання інтернету домінують інші форми залежності. Розуміння типології залежності та її культурних і демографічних детермінант є необхідним для ефективної профілактики.

Сім'я та освітні заклади відіграють ключову роль у ранньому виявленні інтернет залежності. Батьки, за допомогою активного посередництва (обговорення,

встановлення правил), сприяють зменшенню ризику, тоді як надмірно обмежувальні підходи можуть підсилювати компульсивну поведінку. Якість сімейних стосунків значно впливає на звички дітей: міцні зв'язки сприяють здоровому використанню інтернету, а дисфункціональні — провокують надмірну онлайн-активність. Вчителі та шкільні психологи, спостерігаючи за академічною успішністю, соціальною взаємодією та поведінковими змінами, також можуть ідентифікувати проблеми, пов'язані з інтернет залежністю [17].

Комплексні програми цифрової грамотності, що сприяють відповідальному використанню інтернету, а також співпраця між школами, сім'ями та громадськими організаціями, забезпечують узгоджений підхід до вирішення цієї проблеми. Суспільні структури, такі як системи охорони здоров'я, через рутинні скринінги сприяють ранньому виявленню та лікуванню залежності.

Однак культурна нормалізація інтенсивного користування інтернетом ускладнює розмежування між нормативним і патологічним використанням. Це підвищує ризик ігнорування компульсивної поведінки, затримуючи втручання. Для подолання цієї проблеми необхідна зміна парадигми, яка враховує як переваги цифрових технологій, так і їх потенціал спричиняти залежність [18].

Сучасні методи діагностики активно впроваджують аналітику великих даних та штучний інтелект (ШІ), які дозволяють аналізувати великі набори даних для виявлення маркерів інтернет залежності. Використання алгоритмів машинного навчання, таких як згорткові нейронні мережі для обробки ЕЕГ-сигналів, демонструє високу точність у визначенні нейрофізіологічних корелятивів залежності. Аналітика великих даних дозволяє створювати комплексні поведінкові профілі та виявляти аномалії в цифровій поведінці, що служить ранніми індикаторами залежності.

Попри переваги, впровадження ШІ стикається з викликами: конфіденційність даних, етичні питання спостереження та алгоритмічні упередження потребують вирішення. Забезпечення прозорості та інтерпретованості діагностичних рішень є ключовим для зміцнення довіри до цих технологій. Інтеграція ШІ сприяє персоналізованому підходу до виявлення та

управління інтернет залежністю, але вимагає балансу між інноваціями та етикою [19].

1.3. Передумови виникнення залежності серед різних вікових груп

Діти та підлітки особливо вразливі до інтернет залежності через специфічні психологічні та соціальні особливості їхнього розвитку. У підлітковому віці відбуваються значні емоційні та когнітивні зміни, включаючи формування самоідентичності та самооцінки. Потреба в соціальному схваленні стає виразнішою, що спонукає багатьох шукати визнання через соціальні мережі. Ця залежність від онлайн-відгуків часто веде до спотвореного самосприйняття та підвищеної вразливості до адиктивної поведінки.

Надмірне використання цифрових технологій під час критичних періодів когнітивного розвитку асоціюється з кількома негативними наслідками. Тривалий час перед екраном корелює зі зниженням концентрації уваги, порушенням виконавчих функцій та погіршенням академічної успішності. Крім того, постійна взаємодія з цифровими медіа може заважати важливим офлайн-активностям, таким як фізичні вправи та особисте спілкування, які є критичними для здорового розвитку.

Статистичні аналізи виявляють значні відмінності у використанні інтернету серед дітей та підлітків у різних країнах. Дані Міжнародного союзу електрозв'язку показують, що у 2022 році 75% осіб віком від 15 до 24 років у всьому світі були користувачами інтернету, при цьому в розвинених регіонах спостерігалися вищі показники проникнення порівняно з країнами, що розвиваються [20]. Ці відмінності підкреслюють вплив соціально-економічних факторів на цифрову взаємодію та наголошують на необхідності цілеспрямованих втручань для вирішення унікальних проблем, з якими стикається молодь у різних контекстах.

Розглядаючи вразливості дітей та підлітків до інтернет залежності, важливо дослідити існуючі інтервенції, спрямовані на подолання цієї проблеми. Комплексний огляд, проведений Дінг та Лі, надає цінні відомості про сьогоденні тенденції лікування цифрової залежності серед молоді [14]. У своєму методичному аналізі автори ідентифікували 17 рецензованих досліджень, опублікованих у період з 2018 по 2022 роки, які зосереджувалися на інтервенціях, спрямованих на подолання цифрової залежності серед осіб віком від 0 до 18 років.

Огляд показав, що когнітивно-поведінкова терапія (КПТ) та інтервенції на її основі були найбільш поширеними методами. Ці підходи виявилися ефективними у зменшенні симптомів, пов'язаних із цифровою залежністю, таких як тривожність та депресія. КПТ-інтервенції зазвичай спрямовані на допомогу індивідам у розпізнаванні та зміні негативних моделей мислення та поведінки, пов'язаних із використанням інтернету, що сприяє формуванню здоровіших цифрових звичок.

Крім того, автори підкреслили важливість сімейно-орієнтованих інтервенцій. Замість безпосереднього впливу на адиктивну поведінку, ці втручання спрямовані на зміцнення функцій та відносин у родині, що опосередковано сприяє зменшенню цифрової залежності. Створюючи підтримуюче сімейне середовище, діти та підлітки менш схильні вдаватися до надмірного використання інтернету як способу для подолання труднощів.

Огляд також розглянув перспективи цифрових інтервенцій, включаючи веб-платформи, мобільні додатки та технології віртуальної реальності. Ці інноваційні підходи є особливо перспективними, оскільки вони відповідають цифровим уподобанням молоді, роблячи інтервенції більш привабливими та доступними. Наприклад, інтервенції з використанням віртуальної реальності можуть моделювати реальні ситуації, надаючи занурюючий досвід, що сприяє зміні поведінки.

Однак автори критично зазначили кілька обмежень, поширених серед розглянутих досліджень. Серед них: невеликі розміри вибірки, коротка тривалість інтервенцій, відсутність контрольних груп та нерендомізований розподіл учасників. Ці методологічні недоліки ускладнюють узагальнення результатів та

підкреслюють необхідність більш суворих дослідницьких дизайнів у майбутніх дослідженнях.

Підсумовуючи, огляд Дінг та Лі надає всебічний огляд сучасних стратегій втручання для подолання цифрової залежності серед дітей та підлітків. Хоча КПТ та сімейно-орієнтовані інтервенції показали свою ефективність, а цифрові інтервенції мають перспективи, існує чітка потреба у більш надійних емпіричних дослідженнях для розробки та валідації інтервенцій, які можуть бути широко впроваджені для вирішення проблеми цифрової залежності серед молоді.

Інтернет залежність має значні соціальні та освітні наслідки, зокрема погіршення академічної успішності через зниження концентрації, інформаційне перевантаження та порушення сну. Надмірне використання інтернету часто призводить до когнітивної втоми, зменшуючи здатність зосереджуватися на навчальних завданнях і, як наслідок, знижуючи академічні досягнення [21]. Величезна кількість доступної онлайн-інформації може перевантажувати студентів, ускладнюючи обробку та запам'ятовування важливих даних, що ще більше посилює навчальні труднощі. Крім того, тривале використання інтернету часто порушує режим сну, призводячи до недосипання, що негативно впливає на когнітивні функції, необхідні для навчання та консолідації пам'яті [22].

У соціальному аспекті інтернет залежність сприяє ізоляції, зменшуючи кількість особистих контактів і формуючи залежність від онлайн-комунікації. Індивіди можуть відмовлятися від реальних соціальних взаємодій, в результаті чого починається зниження міжособистісних навичок та відчуття самотності. Цей зсув від фізичних до віртуальних взаємодій може послаблювати соціальні зв'язки та зменшувати якість відносин, оскільки онлайн-комунікація часто позбавлена глибини та емоційного насичення, характерних для особистих зустрічей.

Вплив на психічне здоров'я є значним, оскільки інтернет залежність асоціюється з підвищеним рівнем тривожності, депресії та агресії. Компульсивна потреба бути онлайн може викликати стрес і тривогу при обмеженні доступу, тоді як сидячий спосіб життя, пов'язаний з надмірним використанням інтернету, може сприяти розвитку депресивних симптомів. Крім того, вплив онлайн-контенту, що

пропагує агресивну поведінку або кібербулінг, може підвищувати агресивність у вразливих осіб. Ці проблеми психічного здоров'я впливають не лише на особисте благополуччя, але й мають ширші наслідки для соціальних взаємодій та академічної успішності [23].

Отже, інтернет залежність має багатогранні соціальні та освітні наслідки, включаючи погіршення академічної успішності через зниження концентрації, інформаційне перевантаження та порушення сну; соціальну ізоляцію, що виникає внаслідок зменшення реальних взаємодій та залежності від онлайн-комунікації; а також негативні наслідки для психічного здоров'я, такі як тривожність, депресія та агресія. Вирішення цих проблем вимагає комплексних стратегій, що охоплюють освітні інтервенції, підтримку психічного здоров'я та сприяння здоровим звичкам використання інтернету.

Також інтернет залежність має багатогранні економічні наслідки на індивідуальному рівні, що проявляються через прямі фінансові витрати, зниження продуктивності та витрати, пов'язані з лікуванням і реабілітацією. Одним із основних економічних наслідків є надмірні витрати на цифрові послуги, включаючи онлайн-ігри, покупки та різні розважальні платформи. Особи з інтернет залежністю часто вдаються до компульсивних покупок, через що виникають фінансові труднощі і накопичення боргів. Ця модель витрат не лише виснажує особисті заощадження, але й порушує бюджетні плани на основні потреби, тим самим підриваючи загальне економічне благополуччя.

Крім того, інтернет залежність негативно впливає на продуктивність, особливо серед підлітків, які поєднують часткову зайнятість з навчанням. Це зниження продуктивності впливає не лише на поточний дохід, але й перешкоджає розвитку навичок та майбутній працевлаштованості, тим самим накладаючи довгострокові економічні витрати на особу [24].

Окрім прямих фінансових наслідків, інтернет залежність спричиняє непрямі економічні витрати, пов'язані з її лікуванням. Особи можуть потребувати психологічної допомоги, консультування та реабілітаційних послуг для вирішення основних поведінкових проблем. Ці втручання часто вимагають значних

фінансових інвестицій, включаючи оплату терапії, вартість медикаментів та потенційну втрату доходу під час лікування. Більше того, тягар цих витрат може поширюватися на членів сім'ї та установи охорони здоров'я, тим самим посилюючи економічний вплив за межами окремої особи [25].

У сукупності ці фактори підкреслюють багатогранний економічний вплив інтернет залежності на індивідуальному рівні, що вказує на необхідність комплексних стратегій, які охоплюють як профілактику, так і втручання для пом'якшення цих фінансових та продуктивних наслідків.

Інтернет залежність створює багатогранні економічні виклики на макрорівні, охоплюючи зростання витрат для державного та приватного секторів, зниження продуктивності робочої сили та довгострокові втрати у розвитку людського капіталу.

Державні та приватні установи несуть значні витрати на подолання наслідків інтернет залежності. Уряди виділяють ресурси на психологічну підтримку, освітні ініціативи та профілактичні заходи, спрямовані на зменшення поширеності та впливу цього явища. Ці витрати включають фінансування фахівців з психічного здоров'я, розробку інформаційних кампаній та впровадження програм втручання в освітніх системах. Фінансовий тягар також лягає на приватні організації, таких як постачальників медичних послуг та неурядові організації, які інвестують у лікувальні заклади та мережі підтримки для допомоги постраждалим.

Продуктивність робочої сили негативно впливає інтернет залежність, особливо серед працівників, які розвинули звичку надмірного використання інтернету в підлітковому віці. Ця поведінка часто зберігається в дорослому житті, призводячи до зниження концентрації, збільшення прогулів та зменшення загальної ефективності роботи. Кумулятивний ефект зниження продуктивності серед робочої сили призводить до значних економічних втрат через зменшення сукупного обсягу виробництва товарів та послуг. Це явище підкреслює необхідність ранніх втручань для запобігання закріпленню таких звичок, що знижують економічну ефективність [26].

Водночас певні сектори інтернет індустрії отримують економічні вигоди завдяки користувачам із залежною поведінкою. Компанії, що керують соціальними мережами, онлайн-іграми та відеострімінговими сервісами, отримують значні прибутки від збільшення залученості користувачів. Бізнес-моделі цих підприємств часто базуються на тривалому перебуванні користувачів на платформах, використовуючи таргетовану рекламу та внутрішньоігрові покупки для монетизації активності. Хоча це приносить дохід індустрії, виникають етичні питання щодо експлуатації залежностей для фінансової вигоди [27].

У довгостроковій перспективі поширений вплив інтернет залежності становить загрозу для розвитку людського капіталу. Особи, що знаходяться в цьому стані, можуть відчувати затримку у розвитку навичок та зниження соціалізації, що зменшує їхню здатність ефективно сприяти економіці. Втрата основних компетенцій, таких як критичне мислення, комунікація та вирішення проблем, підриває якість робочої сили. Це погіршення людського капіталу має далекосяжні наслідки, потенційно стримуючи інновації, знижуючи конкурентоспроможність та перешкоджаючи сталому економічному зростанню [28].

У сукупності ці фактори висвітлюють складні економічні наслідки інтернет залежності на макрорівні, підкреслюючи необхідність комплексних політичних рамок, що охоплюють профілактику, лікування та етичні аспекти цифрового споживання.

1.4. Методи боротьби з інтернет залежністю: міжнародний досвід

Зростання інтернет залежності стимулювало впровадження національних та міжнародних стратегій для її подолання. У Китаї регуляції обмежують час онлайн-ігор для неповнолітніх, запроваджуючи заборону на доступ у нічний час та систему

реєстрації під справжніми іменами. Ці заходи знизили середній час гри серед молоді, хоча їх ефективність обмежується обхідними шляхами, що потребує доповнення освітніми й психологічними ініціативами [29]. У Південній Кореї впроваджено "Закон про запобігання інтернет залежності" та "Закон про відключення", які поєднують регулятивні заходи з консультаційними й реабілітаційними програмами. Вони зменшили надмірну ігрову поведінку серед підлітків, але викликають дебати щодо особистих свобод та ролі батьківського впливу. Японія, навпаки, обрала освітній підхід, інтегруючи у шкільні програми семінари про ризики надмірного використання інтернету. Ці заходи підвищили обізнаність студентів, однак їхній довгостроковий ефект потребує подальших досліджень [30].

Неурядові організації (НУО) та приватний сектор також сприяють боротьбі з інтернет залежністю. У США НУО, як-от Центр інтернет залежності, впроваджують когнітивно-поведінкову терапію (КПТ) та групову терапію для подолання проблеми, демонструючи покращення психічного здоров'я та навичок саморегуляції. Однак доступність таких програм обмежується фінансовими й географічними бар'єрами. У ЄС приватний сектор розробляє інструменти цифрового добробуту, такі як моніторинг екранного часу та обмеження використання додатків, які підвищують обізнаність користувачів про їхні звички. Хоча ці інструменти сприяють здоровішому цифровому споживанню, їх ефективність залежить від мотивації користувачів і не вирішує основних психологічних причин залежності. Освітні партнерства приватних компаній та шкіл, спрямовані на формування відповідальних цифрових звичок у дітей, демонструють позитивний вплив на зниження ризику інтернет залежності в майбутньому [31].

Зусилля НУО та приватного сектора демонструють важливість багатовимірного підходу до подолання інтернет залежності. НУО зосереджуються на терапевтичних та поведінкових втручаннях, тоді як приватний сектор пропонує технологічні рішення для моніторингу цифрових звичок, разом створюючи

комплексний підхід до вирішення цієї проблеми. Майбутня співпраця між секторами може підвищити ефективність та масштаб таких ініціатив.

Міжнародні ініціативи також відіграють ключову роль у боротьбі з інтернет залежністю. ВООЗ стала лідером цих зусиль, включивши «ігровий розлад» до МКХ-11 у 2018 році. Це визначення через конкретні діагностичні критерії дало наукову основу для розробки національних політик, які сприяють ідентифікації, профілактиці та лікуванню інтернет залежності. Такі країни, як Південна Корея та Китай, вдосконалили свої політики відповідно до рекомендацій ВООЗ, інтегруючи їх у системи охорони психічного здоров'я. Водночас реалізація цих рекомендацій у країнах із низькими доходами залишається складною [32].

ВООЗ також сприяє розвитку потенціалу через методичні посібники, навчальні програми та інформаційні кампанії, підкреслюючи необхідність інтеграції лікування залежності в існуючі системи охорони здоров'я та багатосекторальної співпраці. Міжнародні конференції, як-от Лісабонська конференція з адикцій, виступають платформами для обміну досвідом, демонстрації досліджень та розробки спільних стратегій. Ці події акцентують увагу на культурних аспектах залежності, вимагаючи адаптованих втручань.

Глобальні ініціативи також залучають приватний сектор до розробки інструментів цифрового добробуту, таких як моніторинг екранного часу та алгоритми, що зменшують адиктивну взаємодію. Компанії, як-от Google і Facebook, активно обговорюють ці інструменти на міжнародних форумах, сприяючи їх удосконаленню. Попри це, ефективність міжнародних зусиль залежить від рівня зобов'язань країн та їх здатності адаптуватися до нових викликів, які створює швидка еволюція цифрових технологій.

Таким чином, міжнародна співпраця, ініціативи ВООЗ та приватний сектор формують єдиний глобальний фронт для боротьби з інтернет залежністю. Їхні зусилля сприяють інноваціям, обміну знаннями та створенню здоровіших цифрових звичок у світі.

Сучасні дослідження інтернет залежності є ключовими для розуміння її динаміки, особливо серед молоді, яка є найуразливішою групою. Вони дозволяють

аналізувати психологічні, соціальні та нейрофізіологічні фактори, визначати групи ризику та основні причини, а також розробляти ефективні стратегії профілактики та втручання, як-от когнітивно-поведінкова терапія (КПТ) і прогностні моделі на основі аналізу даних. Ці результати забезпечують науково обґрунтовані підходи для зменшення негативного впливу інтернет залежності.

У статті Девіда Бікхема "Current Research and Viewpoints on Internet Addiction in Adolescents" аналізуються термінологія, рівні поширеності, ризикові чинники та методи лікування. Автор підкреслює нестабільність термінології, що ускладнює порівняння досліджень, та значну варіативність оцінок поширеності залежно від визначень і методів оцінки. Основними ризиковими факторами вважаються імпульсивність, агресивність, нейротизм, а також коморбідні розлади, як-от депресія та СДУГ. Бікхем вказує на ефективність КПТ і фармакологічних втручань у лікуванні залежності, наголошуючи на необхідності стандартизації діагностичних підходів і подальших лонгітюдних досліджень [33].

Дослідження Діотаюті та його колег "Інтернет залежність серед молоді: роль імпульсивності та співзалежності" додає до літератури нові дані про вплив імпульсивності та співзалежності на інтернет залежність. У вибірці з 481 студента університету (18–30 років) за допомогою ієрархічного регресійного аналізу встановлено, що імпульсивність і співзалежність є сильними предикторами залежності, причому імпульсивність виявилася більш значущою [34].

Учасники з вищими рівнями імпульсивності та співзалежності демонстрували значно вираженіші симптоми інтернет залежності. Статеві та вікові особливості також впливали на залежність: чоловіки та старші учасники вибірки частіше мали високі показники залежності, що підкреслює багатогранність цього явища. Дослідження Діотаюті та його колег акцентує увагу на ролі імпульсивності та співзалежності, пропонуючи нове розуміння механізмів формування залежності. Автори наголошують на важливості врахування цих психологічних рис у програмах профілактики та втручання, зокрема для молодих дорослих. На відміну від огляду Бікхема, який зосереджувався на підлітках і досліджував вплив

нейротизму, депресії та дефіциту уваги, це дослідження інтегрує співзалежність як соціальний чинник, що впливає на проблемне використання інтернету.

Методологічно роботи пропонують взаємодоповнюючі перспективи. Бікхем, у своєму огляді, аналізує результати численних досліджень, надаючи широкий погляд на інтернет залежність серед підлітків. Він акцентує увагу на термінологічній нестабільності, коморбідних розладах і поведінкових симптомах. Натомість Діотаюті використовує емпіричний підхід, досліджуючи вплив імпульсивності та співзалежності серед студентів університетів за допомогою стандартизованих оцінок і регресійного аналізу. Такий підхід дозволяє детальніше зрозуміти психологічні риси, що впливають на інтернет залежність у межах конкретної вікової групи.

Досліджувані групи також різняться: Бікхем аналізує підлітків, які є вразливими через тиск однолітків і недостатню саморегуляцію, тоді як Діотаюті досліджує молодих дорослих, які стикаються з академічним стресом, підвищеною автономією та соціальною динамікою в університетському середовищі. Такий порівняльний аналіз виявляє, як вікові особливості та стадії розвитку впливають на прояви інтернет залежності, а також на її ключові чинники.

Попри ці відмінності, обидва дослідження підкреслюють важливість психологічних особливостей у вивченні інтернет залежності. Робота Бікхема пропонує широкий огляд наявних факторів ризику та методологічних викликів, а дослідження Діотаюті детально аналізує взаємодію імпульсивності та співзалежності як ключових предикторів залежності серед молодих дорослих. Результати обох досліджень підкреслюють необхідність адаптації програм профілактики та втручання до потреб різних вікових груп. Для підлітків, як зазначає Бікхем, ефективними можуть бути освітні ініціативи та сімейно-орієнтовані втручання, що сприяють розвитку саморегуляції та зменшенню впливу однолітків. Для молодих дорослих, згідно з Діотаюті, потрібні індивідуалізовані підходи, що враховують їхні психологічні особливості, зокрема імпульсивність і співзалежність. Інтеграція результатів обох досліджень дозволяє глибше зрозуміти,

як індивідуальні риси, вікові характеристики та соціальний контекст взаємодіють у формуванні моделей проблемного використання інтернету.

РОЗДІЛ 2

ЗАСТОСУВАННЯ МЕТОДІВ DATA SCIENCE ДЛЯ АНАЛІЗУ ІНТЕРНЕТ ЗАЛЕЖНОСТЕЙ

2.1. Методика попередньої обробки даних для прогнозування

Прогнозування схильності до інтернет залежності вимагає ретельного та структурованого підходу до обробки, відбору та аналізу даних. Основою будь-якої надійної прогностичної системи є якість та підготовка використовуваних даних, оскільки вони безпосередньо впливають на продуктивність та інтерпретованість моделей. Це особливо важливо в контексті поведінкових явищ, таких як інтернет залежність, де дані часто складаються зі складних, багатовимірних взаємозв'язків, що охоплюють психологічні, поведінкові та демографічні змінні.

У цьому дослідженні попередня обробка даних відіграє ключову роль у забезпеченні того, щоб сирі дані були ефективно трансформовані у формат, придатний для аналізу. Властиві виклики, пов'язані з поведінковими даними, такі як відсутні значення, незбалансовані розподіли та висока розмірність простору ознак, потребують застосування сучасних методів попередньої обробки. Ці методи не лише покращують цілісність та придатність набору даних, але й зменшують потенційні упередження, які можуть негативно вплинути на точність прогнозу моделі.

Етап попередньої обробки включає комбінацію методів вилучення ознак, зменшення розмірності та кодування категоріальних змінних. Ці техніки працюють у синергії, оптимізуючи набір даних для алгоритмів машинного навчання, забезпечуючи включення найрелевантнішої інформації та мінімізацію шуму. Сучасні методи, такі як автоенкодери, застосовуються для виявлення латентних шаблонів у даних, тоді як методи зменшення розмірності допомагають впоратися з

"прокляттям розмірності", дозволяючи моделі зосередитися на найзначущих ознаках. Одночасно методи створення та кодування ознак трансформують сирі дані у формат, придатний для машинного навчання, що дозволяє моделям ефективно інтерпретувати та використовувати різноманітні типи даних.

Цей ретельний підхід до попередньої обробки даних формує основу аналітичної системи, забезпечуючи базу для застосування складних моделей машинного навчання. Зосереджуючись на систематичній трансформації та вдосконаленні даних, ці методи забезпечують надійну основу для подальших етапів моделювання та оцінки.

Агрегація часових рядів є важливою технікою в області науки про дані, особливо при аналізі тимчасових даних для виявлення закономірностей, трендів і сезонних коливань. Цей метод передбачає об'єднання даних, зібраних у часі, в агреговані інтервали, що забезпечує більш зручний і зрозумілий для аналізу набір даних.

Концептуальні основи агрегації часових рядів можна простежити в ранніх статистичних аналізах тимчасових даних. Однак її формалізація та широке застосування з'явилися з розвитком обчислювальної статистики та економетрики в середині XX століття. Дослідники визнали необхідність агрегації даних з високою частотою для зменшення шуму і виділення основних трендів, що призвело до розробки різних методів агрегації.

На практиці агрегація часових рядів використовується для перетворення детальних даних у ширші інтервали, наприклад, перетворення спостережень, зроблених щохвилинно, в годинні, денні або місячні підсумки. Цей процес особливо корисний у випадках, коли дані з високою частотою можуть приховувати довгострокові тренди через короткострокові коливання. Завдяки агрегації даних аналітики можуть зменшити обчислювальну складність і покращити якість візуалізацій, тим самим підвищуючи зрозумілість результатів.

Агрегація часових рядів широко застосовується у багатьох дослідженнях для вирішення конкретних наукових питань. Наприклад, у сфері моделювання енергетичних систем методи агрегації використовувалися для оптимізації

представлення тимчасових змін у постачанні та споживанні енергії. У машинному навчанні агрегація часових рядів інтегрується у моделі прогнозування для підвищення точності прогнозів. Наприклад, дослідження впливу рівнів агрегації на точність прогнозування демонструють баланс між рівнем агрегації та продуктивністю моделі.

У дослідженні "Агрегація часових рядів для оптимізації енергетичних систем: цілі, виклики, підходи та можливості" автори детально аналізують методи агрегації часових рядів, які використовуються для зменшення складності моделей енергетичних систем. Вони підкреслюють, що зростаюча роль відновлюваних джерел енергії вимагає точного представлення часових змінних у моделях оптимізації, оскільки відновлювані джерела характеризуються високою варіативністю та непередбачуваністю. Агрегація часових рядів дозволяє зменшити кількість часових точок, зберігаючи при цьому ключові характеристики даних, що сприяє зниженню обчислювальних витрат та підвищенню ефективності моделей [35].

Автори також обговорюють різні підходи до агрегації, включаючи кластеризацію періодів, вибір типових днів та використання представницьких періодів. Вони зазначають, що вибір методу агрегації повинен враховувати специфіку досліджуваної системи та доступні дані. Крім того, автори рекомендують проводити валідацію отриманих результатів на повному часовому ряді, щоб оцінити точність та надійність агрегованих моделей. Вони підкреслюють важливість балансування між зменшенням розмірності даних та збереженням критичних характеристик, що впливають на результати оптимізації.

Загалом, дослідження демонструє, що агрегація часових рядів є потужним інструментом для підвищення ефективності моделей оптимізації енергетичних систем, особливо в умовах зростаючої частки відновлюваних джерел енергії. Однак успіх застосування цього підходу залежить від ретельного вибору методів агрегації та проведення валідації результатів на повних наборах даних.

У контексті досліджень інтернет залежності агрегація часових рядів може бути важливим інструментом для аналізу моделей поведінки користувачів у

тривалий період. Агрегуючи дані використання в денні або тижневі інтервали, дослідники можуть виявляти тренди та аномалії, які можуть свідчити про проблемну поведінку, що допомагає розробляти стратегії втручання [36].

Агрегація часового ряду забезпечує оптимізацію даних, зберігаючи ключові патерни й зменшуючи обчислювальну складність, що створює міцну основу для подальших етапів аналізу. На наступному етапі важливу роль відіграє Feature Engineering, який дозволяє створювати нові ознаки та трансформувати існуючі, забезпечуючи моделям більш релевантну інформацію для побудови точних прогнозів. Поєднання цих двох процесів підвищує якість підготовки даних, дозволяючи максимально використати їх аналітичний потенціал.

Концепція інженерії ознак розвивалася разом із розвитком машинного навчання та статистичного моделювання. Хоча ранні статистичні методи неявно включали створення ознак через ручні трансформації даних, формалізація інженерії ознак як окремого процесу з'явилася із впровадженням складніших алгоритмів машинного навчання наприкінці XX століття. Дослідники зрозуміли, що якість і релевантність вхідних ознак суттєво впливають на продуктивність моделі, що призвело до систематичного підходу до створення й відбору ознак.

Інженерія ознак охоплює кілька основних завдань. По-перше, створення нових ознак із наявних даних за допомогою математичних перетворень, агрегування або використання знань, специфічних для конкретної галузі. Наприклад, у часових рядах створення лагових змінних або ковзних статистичних показників може надати додаткову інформацію про тимчасові залежності. По-друге, трансформація ознак передбачає їх зміну для відповідності припущенням певних алгоритмів або підвищення їхньої інтерпретованості. Зазвичай застосовуються методи нормалізації, масштабування або перетворення категоріальних змінних у числовий формат. Нарешті, відбір ознак включає ідентифікацію та збереження найбільш релевантних змінних при виключенні зайвих або неінформативних, що є важливим для зменшення розмірності, зниження переобчислення та підвищення ефективності моделі.

Практичне застосування інженерії ознак спостерігається у різних сферах. У сфері обробки природної мови, наприклад, трансформація текстових даних у числові ознаки за допомогою таких методів, як TF-IDF, стала основою завдань класифікації текстів. У фінансовій сфері створення ознак, що відображають ринкові тренди та економічні індикатори, покращило точність прогнозування цін на акції.

Різні дослідження підкреслюють важливість інженерії ознак для підвищення продуктивності моделей. Зокрема, інженерія ознак дозволяє у контексті дослідження інтернет залежності створювати значущі змінні з даних поведінки користувачів, наприклад, тривалості сеансів, частоти активності та моделей використання. Завдяки цьому можна точніше передбачити схильність користувачів до інтернет залежності.

У науковій роботі "Feature Engineering (FE) Tools and Techniques for Better Classification Performance" представлено всебічний огляд методологій, інструментів та технік інженерії ознак, спрямованих на покращення точності класифікаційних моделей [37]. Автори підкреслюють ключову роль інженерії ознак у підвищенні продуктивності моделей машинного навчання, особливо в контексті класифікаційних завдань.

Дослідження охоплює різноманітні підходи до створення та трансформації ознак, включаючи методи відбору ознак, обробку текстових даних, а також специфічні техніки для доменів, таких як клінічна класифікація текстів та прогнозування зв'язків у соціальних мережах. Автори аналізують вплив цих технік на точність моделей, демонструючи, що належно розроблені ознаки можуть суттєво покращити здатність моделей до узагальнення та їхню ефективність на нових даних. Робота підкреслює важливість інженерії ознак як критичного етапу в процесі побудови моделей машинного навчання, що безпосередньо впливає на їхню точність та надійність. Стратегічне використання Інженерії ознак в різних галузях демонструє значне підвищення точності прогнозування, підкреслюючи її цінність як у академічних дослідженнях, так і в практичних застосуваннях.

У сфері науки про дані, особливо при розробці прогнозних моделей, обробка категоріальних змінних є критичним етапом. Кодування категоріальних ознак відноситься до сукупності методів, що використовуються для перетворення категоріальних даних—змінних, які представляють дискретні категорії або мітки—у числові формати, які алгоритми машинного навчання можуть ефективно обробляти. Це перетворення є необхідним, оскільки більшість алгоритмів вимагають числових вхідних даних для обчислення відстаней, кореляцій та інших математичних операцій, що є невід'ємною частиною навчання моделей.

Еволюція методів кодування категоріальних ознак тісно пов'язана з розвитком статистичного моделювання та машинного навчання. Ранні підходи, такі як dummy-кодування та effect-кодування, були розроблені в рамках статистики для включення категоріальних змінних у регресійні моделі. З розвитком технік машинного навчання з'явилися більш складні методи кодування, що відповідають зростаючій складності та масштабам даних [38].

Існує кілька основних типів кодування категоріальних ознак, кожен з яких має свої особливості, переваги та обмеження. Одним із найпростіших методів є Label Encoding, який присвоює унікальне ціле число кожній категорії. Хоча цей підхід є простим, він може ненавмисно вводити порядкові відносини там, де їх немає, що потенційно вводить в оману деякі алгоритми. Для вирішення цієї проблеми було введено One-Hot Encoding, який створює бінарні вектори для кожної категорії, уникаючи нав'язування порядкових відносин. Однак цей метод може призвести до високої розмірності даних, особливо при роботі з ознаками, що мають велику кількість категорій.

Для вирішення проблем високої кардинальності було розроблено Target Encoding. Ця техніка замінює кожну категорію на середнє значення цільової змінної для цієї категорії, ефективно захоплюючи взаємозв'язок між категоріальною ознакою та цільовою змінною. Хоча цей метод є потужним, він може призвести до витоку цільової змінної, якщо не застосовувати його обережно, оскільки використовує інформацію з цільової змінної під час кодування.

Ще одним помітним методом є Binary Encoding, який поєднує властивості як label, так і one-hot кодування. Він перетворює категорії в бінарні числа, а потім розбиває ці бінарні цифри на окремі стовпці. Цей підхід зменшує розмірність порівняно з one-hot кодуванням, зберігаючи унікальність кожної категорії. Однак інтерпретація отриманих закодованих ознак може бути менш інтуїтивно зрозумілою [39].

Вибір методу кодування залежить від різних факторів, включаючи природу категоріальної змінної, використовуваний алгоритм та специфічні характеристики набору даних. Наприклад, алгоритми на основі дерев, такі як дерева рішень та випадкові ліси, зазвичай менш чутливі до вибору кодування, оскільки вони можуть природно обробляти категоріальні змінні. Навпаки, лінійні моделі та алгоритми, що базуються на відстанях, можуть вимагати більш ретельного кодування для забезпечення оптимальної продуктивності.

У контексті прогнозування схильності до інтернет залежності вибір відповідного методу кодування є вирішальним. Такі ознаки, як демографічні дані користувачів, категорії онлайн-активності та поведінкові шаблони, часто мають категоріальний характер. Застосування відповідних технік кодування забезпечує представлення цих ознак у формі, яку прогнознi моделі можуть ефективно використовувати, тим самим підвищуючи точність та надійність прогнозів.

Наступним важливим етапом підготовки даних є методи зменшення розмірності. Це методи, що забезпечує ефективність аналізу у випадках, коли кількість змінних у наборі даних надмірно велика. Після застосування технік створення ознак, які дозволяють отримати нову релевантну інформацію, зменшення розмірності допомагає оптимізувати набір даних, зберігаючи його ключову структуру. Застосування цих методів є критично важливим у контексті великих масивів багатовимірних даних, які можуть містити приховані закономірності та інформаційний шум.

Методи зменшення розмірності спрямовані на зниження кількості змінних у даних без суттєвої втрати інформації. Вперше ці методи почали використовуватися в ранніх статистичних дослідженнях для аналізу багатовимірних даних. Одним із

перших та найпоширеніших методів є аналіз головних компонент (Principal Component Analysis, PCA), який було запропоновано Карлом Пірсоном у 1901 році. PCA перетворює початкові змінні на новий набір некорельованих змінних, які впорядковуються за спаданням дисперсії, яку вони пояснюють. Цей метод є особливо ефективним у випадках, коли дані мають лінійні зв'язки між змінними.

Згодом з'явилися інші підходи, що доповнили методи лінійного зменшення розмірності. Наприклад, лінійний дискримінантний аналіз (Linear Discriminant Analysis, LDA), який було запропоновано Рональдом Фішером у 1936 році, знаходить лінійні комбінації ознак, що забезпечують максимальне розділення між класами. Хоча ці методи є обчислювально ефективними і мають високу інтерпретованість, вони можуть бути обмежені у випадках, коли структура даних є нелінійною.

Для вирішення проблем, пов'язаних із нелінійною структурою даних, було розроблено нелінійні методи зменшення розмірності. Серед таких методів особливу популярність отримали t-SNE (t-distributed Stochastic Neighbor Embedding) та UMAP (Uniform Manifold Approximation and Projection). Метод t-SNE, запропонований Лауренсом ван дер Маатеном та Джеффри Гінтоном у 2008 році, забезпечує візуалізацію високовимірних даних, моделюючи подібні об'єкти у вигляді близько розташованих точок на дво- або тривимірному просторі [40]. UMAP, представлений Ліландом МакІннесом, Джоном Хілі та Джеймсом Мелвіллом у 2018 році, є ще одним інструментом, що дозволяє зберігати як локальні, так і глобальні структури даних, забезпечуючи ефективне зменшення розмірності [41].

Кожен метод має свої переваги та обмеження. Лінійні методи, такі як PCA, характеризуються низькою обчислювальною складністю та простотою інтерпретації, однак вони не здатні враховувати нелінійні залежності. Натомість нелінійні методи, такі як t-SNE та UMAP, дозволяють виявити складні патерни, але вимагають більших обчислювальних ресурсів і можуть бути менш інтерпретованими.

Після застосування методів зменшення розмірності, які дозволяють оптимізувати набір даних, зберігаючи його найважливіші характеристики, наступним кроком є використання більш гнучких і спеціалізованих підходів до роботи з даними. Одним із таких методів є автоенкодера, що поєднують можливості стискання інформації та виявлення прихованих структур у даних. Автоенкодера виходять за межі традиційних методів зменшення розмірності, забезпечуючи як створення ефективного латентного представлення, так і можливість генерації нових даних. Їхня адаптивність і потужність роблять цей підхід важливим інструментом для вирішення складних задач у межах дослідження інтернет залежності.

Автоенкодера — це клас штучних нейронних мереж, розроблених для виконання завдань безконтрольного навчання, зокрема в сфері представлення даних і зменшення розмірності. Їх основна мета полягає у створенні ефективного кодування вхідних даних шляхом навчання мережі реконструювати вхідні дані з їх стислого латентного представлення. Цей процес включає два основні компоненти: енкодер, який перетворює вхідні дані у латентний простір, і декодер, який відновлює дані з цього латентного представлення [42].

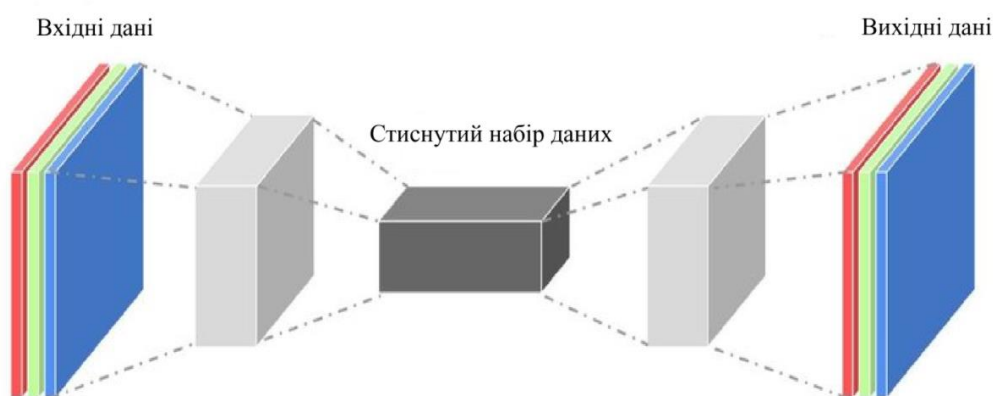


Рисунок 2.1 - Схема архітектури автоенкодера

Джерело: [43]

Концепція автоенкодерів має довгу історію в дослідженнях нейронних мереж. Перші їх реалізації використовувалися для таких завдань, як зменшення розмірності та навчання ознак. З розвитком глибокого навчання автоенкодера

набули особливої популярності в генеративному моделюванні, що дозволяє створювати нові зразки даних, схожі на початковий набір.

Існують різні типи автоенкодерів, які розроблені для вирішення конкретних завдань та застосувань.

Наприклад шумозахищені автоенкодери (Denoising Autoencoders, DAE) є спеціалізованим варіантом автоенкодерів, розробленим для навчання стійких представлень даних шляхом відновлення оригінальних вхідних даних з їх пошкоджених версій. Цей підхід був запропонований у 2008 році для вирішення обмежень традиційних автоенкодерів, які часто не здатні захоплювати змістовні особливості даних, коли навчаються лише на відтворенні вхідних значень.

Основний механізм роботи DAE включає два ключові етапи: цілеспрямоване пошкодження вхідних даних та їх подальше відновлення. Під час навчання кожен зразок даних піддається процесу випадкового пошкодження, наприклад, шляхом додавання гаусівського шуму, маскування випадкових частин вхідних даних або введення шуму типу "сіль і перець". Пошкоджений вхід подається до енкодера, який перетворює його у латентне представлення. Декодер, у свою чергу, намагається відновити оригінальні, непошкоджені дані з цього латентного коду. Мережу навчають мінімізувати розбіжність між оригінальними даними та їх відновленими версіями, що спонукає модель захоплювати приховану структуру даних, а не просто навчатися відтворювати вхідні значення [44].

Цей критерій шумозахисту змушує автоенкодер вивчати представлення, стійкі до порушень у вхідних даних, що дозволяє захоплювати багатовимірну структуру, на якій розташовуються дані. Таким чином, DAE може виявляти внутрішні ознаки, що є інваріантними до шуму, що сприяє створенню більш узагальнених та стійких моделей. Ця властивість є особливо корисною у випадках, коли дані схильні до пошкоджень або при роботі з неповними наборами даних [45].

Ефективність шумозахищених автоенкодерів була продемонстрована у різних галузях. Наприклад, у обробці зображень DAE використовувалися для видалення шуму з пошкоджених зображень, підвищуючи їхню якість. У сфері обробки природної мови вони застосовувалися для відновлення пропущених слів

або виправлення друкарських помилок у текстових даних. Крім того, DAE інтегрувалися у глибокі архітектури машинного навчання для ініціалізації багатoshарових мереж, сприяючи ефективному навчанню моделей завдяки забезпеченню змістовних початкових представлень.

Шумо захищені автоенкодери продемонстрували свою ефективність у навчанні стійких представлень даних завдяки здатності відновлювати пошкоджені вхідні дані. Наступним важливим кроком у розвитку автоенкодерів є розробка їх різновидів, таких як розріджені автоенкодери (Sparse Autoencoders, SAE), які фокусуються на навчанні компактних і значущих представлень через введення обмежень на активацію нейронів у прихованих шарах.

Розріджені автоенкодери — це спеціалізований варіант автоенкодерів, що покликані вивчати ефективні представлення даних за рахунок впровадження обмежень на кількість активованих прихованих нейронів. Такий підхід сприяє тому, що лише частина нейронів у прихованому шарі активується для будь-якого заданого входу, що дозволяє мережі виділяти найбільш важливі ознаки та забезпечувати їхню інтерпретованість.

Архітектура розрідженого автоенкодера подібна до традиційного, енкодер перетворює вхідні дані в латентний простір, а декодер реконструює вхідні дані з цього латентного представлення. Головна відмінність полягає у впровадженні обмеження розрідженості через додатковий штраф у функції втрат. Для цього часто використовують регуляризацию L1 або розрахунок розбіжності Кульбака-Лейблера, що вимірює відхилення середньої активації нейронів від заданого рівня розрідженості.

Таке обмеження на активації стимулює мережу до навчання базисних функцій, що відображають приховану структуру даних. Цей механізм перегукується з концепцією розрідженого кодування в нейронауках, де сенсорна інформація представлена активацією невеликої кількості нейронів. У машинному навчанні такі представлення можуть покращувати продуктивність у завданнях класифікації та кластеризації, оскільки вони виділяють найбільш інформативні ознаки [46].

Розріджені автоенкодери демонструють потужність у навчанні компактних та інтерпретованих представлень даних через введення обмежень на активацію нейронів. Однак, коли постає завдання не лише виявлення ключових особливостей, але й генерації нових зразків даних, що відповідають початковому розподілу, виникає потреба у більш гнучких моделях. У цьому контексті варіаційні автоенкодери стають логічним наступним етапом, забезпечуючи інтеграцію генеративного підходу з можливостями латентного представлення.

Варіаційні автоенкодери (VAE) є складним класом генеративних моделей, які поєднують принципи варіаційного байєсівського висновку з можливостями глибоких нейронних мереж у представленні даних. Вони були запропоновані у 2013 році і з того часу стали важливим інструментом для аналізу складних розподілів даних і генерації нових зразків, які схожі на початковий набір даних. Основною метою VAE є моделювання ймовірнісного розподілу вхідних даних через латентні змінні, які зберігають суттєві особливості інформації. Цей підхід передбачає вибір латентних змінних із заданого апріорного розподілу і подальше генерування спостережуваних даних за допомогою умовного розподілу, що моделює їхню залежність від цих змінних [47].

Проте обчислення точного апостеріорного розподілу є складним завданням через велику кількість параметрів і складність даних. Для подолання цієї проблеми VAE використовують стратегію апроксимації через енкодерну мережу, яка оцінює апостеріорний розподіл за допомогою параметрів, що визначають середнє значення та дисперсію. Ця модель розпізнавання дозволяє отримати оцінку апостеріорного розподілу, що робить навчання моделі можливим.

Процес навчання VAE базується на максимізації нижньої межі доказів, яка слугує метрикою для наближення ймовірності спостережуваних даних. Ця метрика включає два компоненти: втрату реконструкції та регуляризацію. Втрата реконструкції оцінює, наскільки добре модель може відновити вхідні дані із зменшеного представлення, і стимулює створення точних і наближених до оригіналу вихідних даних. Регуляризація, у свою чергу, забезпечує відповідність

між апостеріорним і апіорним розподілами, що дозволяє структурувати латентний простір і забезпечувати безперервність даних.

Однією з ключових складностей у навчанні VAE є необхідність оптимізації стохастичних процесів, які виникають у процесі роботи з латентними змінними. Для вирішення цієї проблеми було впроваджено метод репараметризації, що дозволяє перетворити стохастичні процеси у детерміністичні, що значно полегшує оптимізацію і прискорює навчання.

Варіаційні автоенкодери широко застосовуються для генерації зображень, обробки текстів природної мови, виявлення аномалій та багатьох інших задач. Їхня здатність навчатися змістовним латентним представленням робить їх цінними для задач синтезу даних, аналізу складних розподілів і оцінки невизначеності у результатах. Це забезпечує їхню універсальність і адаптованість до різних реальних сценаріїв. Крім того, ймовірнісна основа VAE забезпечує можливість кількісно оцінювати невизначеність у згенерованих виходах, що розширює їх застосування у реальних сценаріях.

Кожен варіант автоенкодера має свої переваги та обмеження. Шумозахищені автоенкодери підвищують стійкість моделі до шуму в даних, але вимагають ретельного налаштування, щоб уникнути перенавчання. Розріджені автоенкодери сприяють формуванню інтерпретованих ознак, але можуть бути обчислювально складними через введення обмежень на розрідженість. Варіаційні автоенкодери надають строгий підхід до генеративного моделювання, проте можуть створювати зображення низької чіткості, якщо латентний простір не є достатньо регуляризованим.

2.2. Методика прогнозування інтернет залежності на основі методів машинного навчання

Процес побудови прогнозних моделей у науці про дані базується на виборі та впровадженні алгоритмів, які не лише демонструють високу обчислювальну ефективність, але й здатні ефективно адаптуватися до особливостей конкретного набору даних. У контексті прогнозування схильності до інтернет залежності, де дані часто містять складні патерни, численні ознаки та тонкі взаємозалежності, вибір моделей набуває ключового значення. Аналіз моделей, використаних у цьому дослідженні, зумовлений необхідністю не лише точно відображати ці складні взаємозв'язки, але й забезпечувати надійну генералізацію на нових даних.

В цьому дослідженні я застосовував декілька сучасних алгоритмів машинного навчання на основі методів градієнтного бустингу, які були застосовані для досягнення високої прогнозової точності у даному дослідженні. Ці моделі, обрані через їхню доведену ефективність у роботі зі структурованими даними, особливо успішно виявляють нелінійні залежності та враховують взаємодію між ознаками, що є критично важливим для поведінкових наборів даних.

Градiєнтний бустинг є потужним і універсальним методом ансамблевого навчання, який став одним із основних інструментів сучасного прогнозного моделювання, особливо для задач, що стосуються структурованих даних. Цей метод був вперше запропонований у кінці 1990-х років Джеромом Г. Фрідманом, який розширив концепцію бустингу, інтегрувавши принципи градієнтної оптимізації. Сам бустинг, що виник раніше завдяки алгоритму AdaBoost, базувався на комбiнуванні слабких моделей в потужний ансамбль шляхом ітеративного коригування їхніх ваг на основі рівня помилок класифікації. Градієнтний бустинг удосконалив цю концепцію, використовуючи градієнтний спуск для мінімізації диференційованої функції втрат, що дозволило поширити підхід на задачі регресії та класифікації [48].

Основою градієнтного бустингу є його ітеративна та адитивна природа. Він починається з простої моделі, зазвичай одного дерева рішень, яке виступає як слабкий учень і забезпечує початкові прогнози. Наступні моделі поступово тренуються для того, щоб зосередитися на корекції залишкових помилок — різниці між фактичними та прогнозованими значеннями, отриманими попередніми

моделями. Цей поетапний процес уточнення, керований градієнтами функції втрат щодо прогнозів, дозволяє ансамблю поступово зменшувати помилки. На кожній ітерації модель наближається до реального розподілу даних, оптимізуючи точність прогнозів і одночасно запобігаючи перенавчанню завдяки таким технікам, як зменшення швидкості навчання (learning rate) і рання зупинка (early stopping).

Ключовим компонентом градієнтного бустингу є дерево рішень, яке виконує роль базового учня в цьому ансамблевому підході. Дерева рішень — це ієрархічні моделі, які розділяють дані на підмножини на основі значень ознак, використовуючи умови, визначені у вузлах. Ці розділення спрямовані на максимізацію однорідності у створених гілках, ефективно сегментуючи дані на регіони зі схожими цільовими значеннями. Листя дерева відповідають прогнозам, тоді як внутрішні вузли визначають правила ухвалення рішень, що базуються на ознаках. Дерева рішень особливо підходять для цієї ролі завдяки своїй здатності працювати з нелінійними залежностями та взаємодією між ознаками, а також адаптуватися до різних розподілів даних [49].

Структура дерев рішень робить їх інтерпретованими та обчислювально ефективними, проте їхня індивідуальна прогностична здатність відносно слабка, що відповідає вимогам до слабких моделей у бустинговій архітектурі. Градієнтний бустинг використовує цю слабкість, поєднуючи виходи багатьох таких дерев, кожне з яких поступово вносить свій внесок у загальну модель. Завдяки фокусуванню на залишкових помилках на кожній ітерації ансамбль у цілому навчається вирішувати складні задачі, які одне дерево не може охопити. Цей процес трансформує набір слабких учнів у високоточну і генералізовану модель.

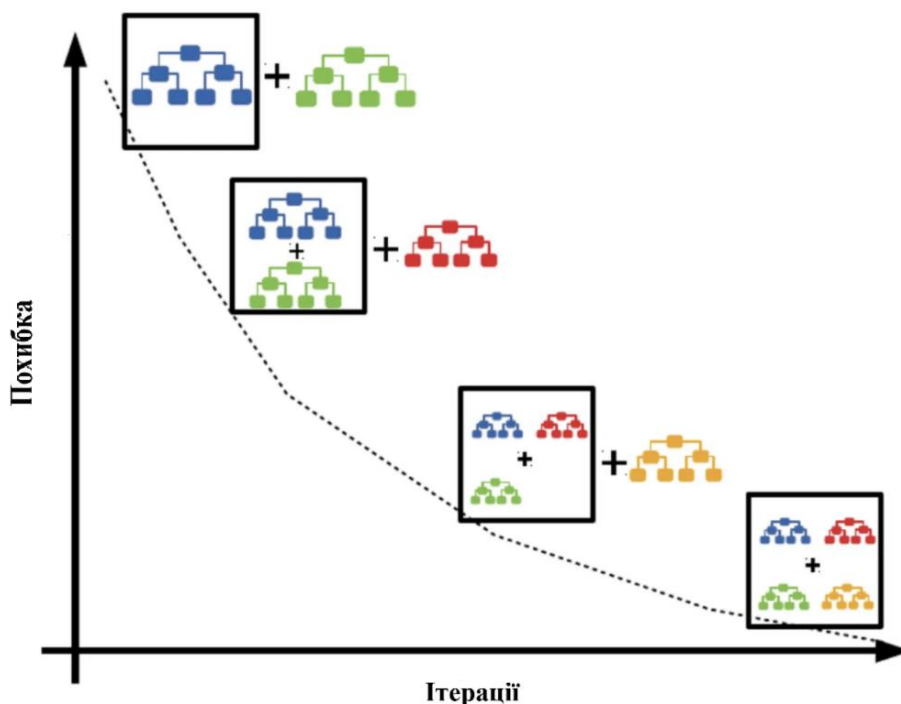


Рисунок 2.2 - Схема навчання алгоритмів градієнтного бустингу

Джерело: [50]

Градієнтний бустинг здобув значну увагу завдяки своїй ефективності у роботі з наборами даних, які містять складні взаємозв'язки між ознаками та нелінійні патерни. Його здатність адаптуватися до різних функцій втрат, включаючи середньоквадратичну помилку для регресії та логарифмічну втрату для класифікації, підкреслює його універсальність. Крім того, його можливість включати регуляризаційні техніки, такі як штрафи L1 та L2, підвищує його стійкість, запобігаючи перенавчанню навіть при роботі з високовимірними або шумними даними.

Історично впровадження градієнтного бустингу стало переломним моментом у застосуванні ансамблевих методів, розширивши їхню корисність далеко за межі обмежень традиційних підходів, таких як бэггінг або простий бустинг. Завдяки цьому градієнтний бустинг став незамінним інструментом у науці про дані та машинному навчанні, забезпечуючи фундамент для багатьох сучасних алгоритмів та реалізацій, які продовжують розвиватися і вдосконалювати можливості прогнозного моделювання.

Алгоритм бустингу функціонує через ітеративний процес, який підвищує точність прогнозування шляхом послідовного усунення недоліків попередніх моделей. Він починається з простої базової моделі, часто званої слабким учнем, яка надає початкові оцінки цільової змінної. Наступні моделі навчаються на залишкових помилках — різниці між фактичними значеннями цільової змінної та прогнозами, згенерованими попередньою моделлю. Зосереджуючись на цих залишках, кожна нова модель прагне виправити помилки свого попередника, поступово підвищуючи загальну точність прогнозування ансамблю.

Ключовим параметром у цьому процесі є швидкість навчання, яка визначає ступінь впливу кожної наступної моделі на сукупний прогноз. Низька швидкість навчання призначає меншу вагу кожній новій моделі, що вимагає більшої кількості ітерацій для досягнення збіжності, але часто призводить до більш стійкої моделі, менш схильної до перенавчання. Навпаки, висока швидкість навчання прискорює збіжність, але може призвести до перенавчання, оскільки модель може надто точно підлаштовуватися під навчальні дані, сприймаючи шум як значущу закономірність. Отже, вибір відповідної швидкості навчання є важливим для балансування між швидкістю збіжності та узагальненням моделі.

Техніки регуляризації відіграють вирішальну роль у запобіганні перенавчанню в рамках бустингу. Методи, такі як зменшення швидкості навчання, що передбачає застосування швидкості навчання менше одиниці, та субсемплінг, де для навчання кожної моделі використовується випадкова підмножина даних, допомагають контролювати складність моделі та підвищувати її здатність до узагальнення. Крім того, обмеження на глибину окремих дерев рішень, кількість листків або мінімальну кількість зразків, необхідних для розділення вузла, застосовуються для запобігання надмірній складності моделі та перенавчанню на навчальних даних.

Взаємозв'язок між кількістю ітерацій бустингу та складністю базових учнів, зазвичай дерев рішень, є важливим аспектом алгоритму бустингу. Збільшення кількості ітерацій дозволяє ансамблю захоплювати більш складні закономірності в даних, але також підвищує ризик перенавчання, особливо якщо базові учні є

складними. З іншого боку, використання простіших базових учнів може вимагати більше ітерацій для досягнення задовільної продуктивності, але може призвести до більш узагальненої моделі. Таким чином, необхідно знайти обережний баланс для оптимізації як зміщення, так і дисперсії компонентів помилки моделі.

Здатність алгоритму бустингу зменшувати як зміщення, так і дисперсію впливає з його ітеративного підходу до вдосконалення моделі. Послідовно зосереджуючись на залишкових помилках попередніх моделей, бустинг ефективно зменшує зміщення, оскільки кожна нова модель виправляє помилки своїх попередників. Одночасно, завдяки технікам регуляризації та ретельному управлінню складністю моделі, бустинг контролює дисперсію, запобігаючи підлаштуванню моделі під шум у навчальних даних. Ця подвійна здатність дозволяє бустингу адаптуватися до складних наборів даних із взаємозалежними ознаками, захоплюючи тонкі взаємозв'язки, які можуть бути пропущені іншими підходами до моделювання.

Серед найпоширеніших методів градієнтного бустингу виділяються LGBMRegressor, CatBoostRegressor та XGBRegressor, кожен з яких характеризується унікальними теоретичними основами та практичними реалізаціями.

LGBMRegressor, реалізація Light Gradient Boosting Machine (LightGBM), використовує підхід навчання дерев рішень на основі гістограм. Ця техніка дискретизує безперервні ознаки в дискретні бінарні значення, зменшуючи кількість точок розбиття, що розглядаються під час побудови дерева. Така дискретизація не лише прискорює процес навчання, але й підвищує ефективність використання пам'яті, що робить LGBMRegressor особливо придатним для роботи з великими наборами даних із високою розмірністю ознак. Стратегія зростання дерев за принципом "leaf-wise" (найкраще спочатку) додатково сприяє його обчислювальній ефективності, оскільки дозволяє глибше досліджувати найбільш значущі гілки, що призводить до точнішого моделювання складних закономірностей у даних.

З іншого боку, CatBoostRegressor розроблений для ефективної обробки категоріальних ознак без необхідності масштабної попередньої обробки. Він впроваджує інноваційну техніку, відому як "ordered boosting", яка зменшує ризик витоку цільової змінної, забезпечуючи, що інформація, використана для кодування категоріальних змінних, походить виключно з попередніх точок даних. Цей метод зберігає цілісність процесу навчання та підвищує здатність моделі до узагальнення. Крім того, CatBoostRegressor будує симетричні дерева, де кожен рівень дерева збалансований. Ця симетрія сприяє стабільності прогнозів і є особливо корисною при роботі з незбалансованими наборами даних, оскільки запобігає упередженості моделі до більш поширених класів.

XGBRegressor, що представляє фреймворк eXtreme Gradient Boosting, відомий своєю надійністю та гнучкістю. Він використовує складний механізм регуляризації, який штрафує за складність моделі, запобігаючи перенавчанню та сприяючи узагальненню. XGBRegressor також підтримує паралельну обробку, що прискорює фазу навчання, особливо при роботі з великими обсягами даних. Його здатність внутрішньо обробляти відсутні значення та підтримка користувацьких функцій втрат додатково підвищують його адаптивність у різних завданнях прогнозування.

Порівнюючи ці алгоритми, виникають кілька компромісів щодо обчислювальної ефективності, точності, споживання пам'яті та продуктивності на різних типах даних. Гістограмний підхід LGBMRegressor та стратегія зростання дерев "leaf-wise" забезпечують високу швидкість обчислень та ефективність використання пам'яті, що робить його придатним для великих наборів даних. Однак цей метод може вимагати ретельного налаштування для запобігання перенавчанню, особливо в умовах обмежених даних. Здатність CatBoostRegressor обробляти категоріальні ознаки без масштабної попередньої обробки спрощує процес моделювання та зменшує ймовірність помилок, спричинених людським фактором, під час підготовки даних. Його симетрична структура дерев підвищує стабільність прогнозів, що є корисним для незбалансованих наборів даних. Проте обчислювальні витрати, пов'язані з підтриманням симетрії, можуть призвести до

довшого часу навчання порівняно з LGBMRegressor. XGBRegressor пропонує збалансований підхід завдяки своїм надійним технікам регуляризації та підтримці паралельної обробки, забезпечуючи компроміс між обчислювальною ефективністю та точністю моделі. Його гнучкість у роботі з різними типами даних та підтримка користувацьких функцій втрачають його універсальним вибором для широкого спектра завдань прогнозного моделювання [51].

У сфері прогнозного моделювання, інтеграція кількох регресійних алгоритмів через методи ансамблю продемонструвала значне підвищення точності прогнозування та стійкості моделей. Вибірковий регресійний ансамбль функціонує як метаалгоритм, що об'єднує результати кількох моделей, таких як LGBMRegressor, XGBRegressor та CatBoostRegressor, для створення комбінованого прогнозу, який перевищує продуктивність окремих компонентів.

Механізм роботи вибіркового регресійного ансамблю передбачає синтез прогнозів його складових моделей, що може здійснюватися за допомогою зваженого або незваженого усереднення. У разі незваженого усереднення кожен внесок моделей у кінцевий прогноз розглядається рівнозначно, що дозволяє мінімізувати індивідуальні похибки й упередження. Зважене усереднення, натомість, призначає різні ваги моделям на основі їхньої точності прогнозування, яка зазвичай визначається через крос-валідацію або інші метрики продуктивності, тим самим підкреслюючи вплив більш точних моделей на вихідні результати ансамблю [52].

Однією з основних переваг використання вибіркового регресійного ансамблю є його підвищена стійкість до перенавчання. У той час як окремі моделі можуть перенавчатися на конкретні патерни або шум у навчальних даних, процес агрегації в ансамблі зазвичай згладжує такі особливості, що забезпечує покращене узагальнення для нових даних. Ця властивість є особливо корисною в складних наборах даних, де ризик перенавчання значно зростає.

Емпіричні дослідження підтверджують, що у сценаріях з високою розмірністю або неоднорідністю даних вибіркові регресійні ансамблі перевершують окремі моделі. Наприклад, у наборах даних, де окремі моделі

можуть стикатися зі складнощами у всебічному виявленні підлеглої структури, ансамблевий підхід, поєднуючи різні перспективи від своїх складових моделей, ефективніше моделює складні залежності та взаємодії у даних. Це сприяє підвищенню точності прогнозів та глибшому розумінню факторів, які впливають на схильність до інтернет залежності [53].

2.3. Методи вдосконалення навчання моделей і оцінки їхньої ефективності

У прогнозній аналітиці вдосконалення методів навчання та оцінки моделей є критичним для отримання надійних і точних результатів. Використання вдосконалених технік навчання дозволяє виявляти приховані закономірності в даних, мінімізуючи ризики перенавчання та недонавчання. Перенавчання виникає через засвоєння шуму в даних, що знижує генералізацію, тоді як недонавчання є наслідком недостатньої складності моделі для відображення даних. Для підвищення генералізації застосовуються регуляризація, оптимізація гіперпараметрів, вибір відповідної архітектури моделей та ансамблеві методи, які поєднують прогнози кількох моделей, підвищуючи точність і стійкість [54].

Ефективна оцінка моделей передбачає використання крос-валідації для перевірки продуктивності на різних підмножинах даних. Зокрема, k -fold крос-валідація, де набір даних ділиться на k фолдів, забезпечує всебічну оцінку, усереднюючи метрики продуктивності. Стратифікована k -fold крос-валідація, яка зберігає пропорції класів у кожному фолді, є особливо корисною для незбалансованих даних, як у випадках прогнозування схильності до інтернет залежності, забезпечуючи надійні та неупереджені оцінки [55].

Крос-валідація запобігає перенавчанню, оцінюючи варіативність і стабільність моделі. Узгоджені результати між фолдами підтверджують здатність моделі узагальнювати на нові дані. Однак, для задач із неоднорідними даними,

такими як прогнозування схильності до інтернет залежності, важливим етапом є оптимізація порогу прийняття рішень. Налаштування порогу ймовірності дозволяє збалансувати чутливість (істинно позитивні результати) і специфічність (істинно негативні результати), що особливо важливо для мінімізації наслідків хибнопозитивних і хибнонегативних прогнозів [56].

Стандартний поріг 0,5 у класифікаційних моделях часто не є оптимальним для незбалансованих даних, оскільки сприяє прогнозуванню на користь більшості класу. Коригування порогу дозволяє досягти балансу між чутливістю та специфічністю, залежно від цілей завдання. У прогнозуванні схильності до інтернет залежності можна акцентувати на підвищенні чутливості для ідентифікації осіб у зоні ризику або на специфічності для зменшення хибнопозитивних результатів [57].

Методи оптимізації порогу, такі як аналіз кривих ROC або Precision-Recall, забезпечують вибір значення, яке найкраще відображає баланс між метриками. Це особливо важливо для незбалансованих даних, де традиційні метрики, як точність, можуть вводити в оману. При прогнозуванні інтернет залежності набори даних зазвичай включають числові та категоріальні ознаки із нерівномірним розподілом класів. Моделі градієнтного бустингу, такі як LGBM, XGB і CatBoost, адаптовані до цих умов завдяки здатності обробляти складні взаємозв'язки між ознаками та незбалансовані дані. CatBoost ефективно працює з категоріальними даними, LGBM оптимізує обчислення через гістограмний підхід, а XGB забезпечує регуляризацію для зменшення перенавчання.

Стратифікована k-fold крос-валідація підходить для таких моделей, оскільки зберігає розподіл класів у кожному фолді, забезпечуючи неупереджені оцінки. Це критично важливо для прогнозування інтернет залежності, де класи меншості часто є ключовими для результатів. Додатково оптимізація порогів дозволяє налаштувати моделі для досягнення бажаного балансу між чутливістю та специфічністю, що важливо для задач із різними наслідками хибнопозитивних і хибнонегативних результатів. Інтеграція цих методів і моделей забезпечує точність, стійкість і надійність прогнозів.

Квадратичний зважений каппа (QWK) є статистичною метрикою, яка враховує ступінь невідповідності між передбачуваними та фактичними класами, особливо актуальною у порядкових класифікаціях. QWK вимірює відповідність через зважену матрицю плутанини, де кожен елемент штрафує за неправильну класифікацію залежно від квадратичної різниці між індексами передбаченого та фактичного класів. Значення QWK варіюється від -1 (гірше за випадковість) до 1 (ідеальна відповідність), із 0, що відображає відповідність на рівні випадковості. Процес розрахунку QWK включає створення матриці плутанини, обчислення спостереженої відповідності та порівняння її з очікуваною відповідністю, яка базується на граничних розподілах класів. Ця метрика дозволяє моделювати тяжкість помилок, надаючи менші штрафи прогнозам, які близькі до істини, і більші — за значні відхилення. Це робить QWK більш інформативною, ніж традиційні метрики, як-то точність, особливо для задач із порядковою природою класів, таких як прогнозування поведінкових ризиків.

Моделі градієнтного бустингу, як-от LGBM, XGB і CatBoost, ефективно інтегруються з QWK, забезпечуючи точність і стійкість у порядкових задачах. Завдяки здатності цих моделей обробляти змішані та незбалансовані дані, вони добре підходять для задач прогнозування інтернет залежності, де класи меншості мають особливе значення. Стратифікована k-fold крос-валідація додатково підтримує оцінку моделей, зберігаючи розподіл класів і забезпечуючи точнішу оцінку продуктивності. Такий підхід дозволяє не лише досягти високої точності, а й мінімізувати критичні помилки, що підвищує надійність моделей у реальних сценаріях.

РОЗДІЛ 3

ПРОГНОЗУВАННЯ ІНТЕРНЕТ ЗАЛЕЖНОСТІ МЕТОДАМИ МАШИННОГО НАВЧАННЯ

3.1. Характеристика даних: структура та джерела

Цей набір даних походить із Healthy Brain Network (HBN), масштабної ініціативи у сфері психічного здоров'я, спрямованої на виявлення біологічних та поведінкових маркерів для вдосконалення діагностики та лікування психічних і навчальних розладів [58]. Набір даних включає клінічну вибірку приблизно 5000 учасників віком від 5 до 22 років, які пройшли детальні клінічні та дослідницькі обстеження. Зокрема, цей набір даних слугує основою для розробки моделей прогнозування, які виявляють проблемне використання інтернету на основі даних про фізичну активність та поведінкових показників.

Набір даних, містить табличні дані, отримані з різноманітних вимірювальних інструментів. Інструменти, що використовуються для збору даних, включають:

1. Демографічні показники – містять інформацію про вік і стать учасників.
2. Використання інтернету – показник кількості годин щоденного використання комп'ютера або інтернету.
3. Глобальна шкала оцінки дітей (Children's Global Assessment Scale) – числова шкала, що використовується фахівцями з психічного здоров'я для оцінки загального функціонування дітей віком до 18 років.
4. Фізичні показники – включають дані про артеріальний тиск, частоту серцевих скорочень, зріст, вагу, а також об'єми талії та стегон.
5. Кардіоваскулярна фізична підготовка (FitnessGram Vitals and Treadmill) – вимірювання стану серцево-судинної системи за допомогою протоколу бігової доріжки NHANES.

6. Фізична підготовка дітей (FitnessGram Child) – оцінка фізичної підготовки за п'ятьма параметрами: аеробна здатність, м'язова сила, м'язова витривалість, гнучкість і склад тіла.
7. Біоелектричний імпедансний аналіз (Bio-electric Impedance Analysis) – вимірювання ключових компонентів складу тіла, включаючи індекс маси тіла (BMI), вміст жиру, м'язів і води.
8. Анкета фізичної активності (Physical Activity Questionnaire) – дані про участь дітей у активних фізичних вправах протягом останніх 7 днів.
9. Шкала порушень сну (Sleep Disturbance Scale) – шкала, що класифікує порушення сну у дітей.
10. Актиграфія (Actigraphy) – об'єктивне вимірювання фізичної активності в природному середовищі за допомогою дослідницького біотрекера.
11. Тест інтернет залежності "Батько-Дитина" (Parent-Child Internet Addiction Test, PCIAT) – 20-пунктова шкала для вимірювання характеристик і поведінки, пов'язаних із компульсивним використанням інтернету, включаючи компульсивність, ескапізм і залежність.

Δ Id	Δ Basic_De...	# Basic_De...	# Basic_De...	Δ CGAS-Se...	# CGAS-CG...	Δ Physical-...	# Physical-...	# Physical-...	# Physical-...
00008ff9	Fall	5	0	Winter	51	Fall	16.87731569	46.0	50.8
000fd460	Summer	9	0			Fall	14.03559028	48.0	46.0
00105258	Summer	10	1	Fall	71	Fall	16.64869606	56.5	75.6
00115b9f	Winter	9	0	Fall	71	Summer	18.29234694	56.0	81.6
0016bb22	Spring	18	1	Summer					
001f3379	Spring	13	1	Winter	50	Summer	22.27995198	59.5	112.2

Рисунок 3.1 - Приклад даних про учасників

Джерело: розроблено автором

Особливе значення має поле PCIAT_Total, яке використовується для формування цільової змінної SII.

Індекс тяжкості порушень (Severity Impairment Index, SII) є стандартизованим показником, що використовується для кількісної оцінки рівня проблемного використання інтернету. Він відображає ступінь впливу інтернет

залежності на повсякденне життя індивіда, включаючи соціальні, академічні та професійні аспекти.

Розрахунок SII базується на результатах спеціалізованих опитувальників, які містять серію питань, спрямованих на виявлення симптомів інтернет залежності. Одним із найпоширеніших інструментів є Тест на інтернет залежність (Internet Addiction Test, IAT), розроблений Кімберлі Янг. Цей тест складається з 20 питань, кожне з яких оцінюється за шкалою від 0 до 5 балів, де 0 означає "ніколи", а 5 — "завжди". Сумарний бал, отриманий за результатами тесту, може варіюватися від 0 до 100 і інтерпретується наступним чином:

- 0–30 балів — нормальне використання інтернету;
- 31–49 балів — легкий ступінь інтернет залежності;
- 50–79 балів — помірний ступінь інтернет залежності;
- 80–100 балів — надмірне користування інтернетом.

Таким чином, SII відображає загальний бал, отриманий за результатами IAT, і використовується для визначення ступеня інтернет залежності у досліджуваних осіб.

Дані актиграфії, отримані з акселерометрів, надають детальну інформацію про фізичну активність учасників, включаючи інтенсивність і тривалість рухів. Ці дані є цінними для вивчення кореляцій між рівнем фізичної активності та поведінкою в інтернеті. З іншого боку, табличні дані забезпечують ширший контекст, включаючи самозвіти про звички, демографічні фактори та показники фізичної підготовки.

Набір даних з актиграфії містить часові ряди, отримані за допомогою акселерометрів, які учасники дослідження HBN носили під час виконання повсякденних справ. Дані доступні у форматі parquet, розбиті за унікальним ідентифікатором id. Основні характеристики набору даних:

1. id – Унікальний ідентифікатор учасника.
2. step – Цілочисельний показник часу, що представляє крок спостереження в межах серії.

3. X, Y, Z – Вимірювання прискорення (у g), зафіксовані акселерометром уздовж стандартних осей.
4. enmo – Обчислена за допомогою пакету *wristpy* метрика Euclidean Norm Minus One (ENMO) для сигналів акселерометра (x-, y-, z-осі, вимірюваних у g). Негативні значення округлюються до нуля; нульові значення вказують на періоди відсутності руху.
5. anglez – Кут нахилу руки відносно горизонтальної площини, обчислений за допомогою *wristpy*.
6. non-wear_flag – Бінарний індикатор (0 – акселерометр носить, 1 – акселерометр знято), визначений за стандартами GGIR на основі стандартного відхилення та діапазону даних акселерометра.
7. light – Вимірювання освітленості в люксах.
8. battery_voltage – Напруга батареї акселерометра, виміряна в мілівольтах (mV).
9. time_of_day – Час доби, що відповідає початку 5-секундного інтервалу, в якому було зібрано дані; формат представлення – %H:%M:%S.%9f.
10. weekday – День тижня, закодований як ціле число (1 – понеділок, 7 – неділя).
11. quarter – Чверть року, представлена цілим числом від 1 до 4.
12. relative_date_PCIAT – Кількість днів (ціле число), що минули з моменту проведення тесту PCIAT (від’ємні значення вказують на дані, зібрані до тесту).

	step	X	Y	Z	enmo	anglez	non-wear_flag	light	battery_voltage	time_of_day	weekday	quarter	relative_date_PCIAT
0	0	0.056732	0.296526	-0.701440	0.108617	-56.464138	0.0	36.000000	4180.000000	42120000000000	1	4	45.0
1	1	-0.062789	0.043119	-0.991394	0.000091	-85.607712	0.0	41.666668	4179.666504	42125000000000	1	4	45.0
2	2	-0.055851	0.042832	-0.992437	0.005979	-85.936378	0.0	29.000000	4179.333496	42130000000000	1	4	45.0
3	3	-0.046101	0.045321	-0.992723	0.000103	-86.287697	0.0	21.000000	4179.000000	42135000000000	1	4	45.0
4	4	-0.060056	0.040616	-0.991980	0.001455	-85.713974	0.0	17.500000	4178.666504	42140000000000	1	4	45.0

Рисунок 3.2 - Приклад даних з акселерометрів

Джерело: розроблено автором

Healthy Brain Network, очолювана Child Mind Institute у співпраці з лідерами громад і сім'ями, надає унікальне сховище даних, що поєднує клінічну точність із актуальністю для реального світу. У цьому дослідженні використовується набір

даних, який включає показники фізичної активності, отримані за допомогою наручних акселерометрів (актиграфія), оцінки фізичної підготовки та опитувальники. Ці фізичні показники, які легко доступні та потребують мінімального втручання, слугують проксі-показниками для прогнозування поведінки в інтернеті. Метою є зв'язок між фізичною активністю та Індексом тяжкості порушень (SII), який є перевіреною метрикою для оцінки серйозності проблемного використання інтернету.

Використання фізичних показників як проксі-показників для поведінки в інтернеті дозволяє подолати значні проблеми сучасного діагностичного ландшафту. Традиційні методи оцінки проблемного використання інтернету часто покладаються на складні клінічні обстеження, створюючи бар'єри, такі як висока вартість, культурна невідповідність і обмежений доступ у малозабезпечених регіонах. Завдяки використанню показників фізичної активності це дослідження прагне надати альтернативний діагностичний підхід, який є масштабованим і практичним у різних умовах, дозволяючи раніше виявляти та втручатися у проблеми.

Основний файл, який слугує ядром дослідження, містить інформацію про 3960 учасників, кожен із яких представлений одним спостереженням. Ця таблиця включає 82 змінні, що охоплюють широкий спектр демографічних, фізичних, психологічних і поведінкових характеристик. Основною метою цього набору даних є забезпечення багатовимірного підходу до аналізу факторів, що впливають на проблемне використання інтернету, через поєднання кількісних та якісних індикаторів. Усі дані були отримані із науково перевірених інструментів і методик, забезпечуючи високий рівень достовірності.

Ключовими змінними в наборі даних є вік та стать учасників, які забезпечують базові демографічні параметри. Вік учасників представлений у вигляді числового значення і є одним із головних факторів для аналізу, оскільки рівень інтернет залежності може суттєво варіюватися залежно від вікових категорій. Стать, зафіксована як категоріальна змінна, дозволяє досліджувати гендерні аспекти проблемного використання інтернету.

Серед фізичних змінних у наборі даних виділяються параметри, отримані з біометричних вимірювань, такі як зріст, вага, індекс маси тіла (BMI), артеріальний тиск (систолический та діастолічний) і частота серцевих скорочень. Ці показники забезпечують комплексну оцінку фізичного стану учасників, що є важливим для вивчення зв'язку між фізичною активністю та поведінковими звичками. Дані, зібрані за допомогою біоелектричного імпедансного аналізу (BIA), включають такі параметри, як загальний вміст жиру, м'язова маса, вода в організмі, базальна швидкість метаболізму та інші елементи, що дозволяють оцінювати склад тіла з високою точністю.

Дані щодо фізичної активності представлені результатами анкетування PAQ-A та PAQ-C, які відображають рівень участі дітей у фізичних активностях протягом останнього тижня. Крім того, показники фізичної підготовки, такі як результати тесту FitnessGram та параметри кардіоваскулярної витривалості, оцінені за допомогою протоколу бігової доріжки NHANES, доповнюють загальну картину фізичного здоров'я.

Одним із центральних компонентів цього файлу є результати опитування Parent-Child Internet Addiction Test (PCIAT), який складається з 20 пунктів і вимірює рівень інтернет залежності через оцінку таких характеристик, як компульсивність, уникнення та залежність. Загальна оцінка PCIAT_Total виступає основою для розрахунку індексу тяжкості порушень (SII), який є ключовим показником у цьому дослідженні. Значення SII визначається за шкалою від 0 до 3, де 0 вказує на відсутність проблемного використання інтернету, а 3 — на його серйозний рівень.

Додатково до фізичних та психологічних змінних у файлі містяться дані про години використання інтернету та комп'ютера, що дозволяє оцінити поведінкові патерни учасників. Інші психологічні інструменти, такі як Шкала оцінки глобального функціонування дітей (CGAS) та Шкала порушення сну (SDS), надають цінну інформацію про загальний психічний стан учасників.

Ця багатовимірна структура даних дозволяє проводити комплексний аналіз зв'язків між фізичними, психологічними та поведінковими характеристиками

учасників, відкриваючи можливості для виявлення причинно-наслідкових зв'язків і побудови прогностичних моделей високої точності.

Файли Parquet, які містять дані з акселерометрів, є важливим компонентом цього дослідження. Вони забезпечують часові ряди, отримані з пристроїв, які учасники носили на зап'ясті протягом до 30 днів. Ці дані надають можливість досліджувати фізичну активність учасників у звичному для них середовищі. Вони охоплюють 13 змінних, кожна з яких надає окремий аспект інформації про активність, стан пристрою чи часовий контекст.

Змінна `id` є унікальним ідентифікатором кожного учасника, що дозволяє співвідносити дані акселерометрії з іншими змінними в основному файлі CSV. Вона забезпечує інтеграцію різних джерел даних для кожного учасника. Змінна `step` використовується для позначення кожного запису у серії акселерометричних даних, маючи свій часовий індекс, представлений у вигляді цілого числа. Вона допомагає впорядкувати спостереження у часових рядах та аналізувати динаміку фізичної активності протягом часу.

Дані прискорення представлені змінними `X`, `Y` та `Z`, які відображають вимірювання прискорення в одиницях g (гравітаційної сили) вздовж трьох стандартних координатних осей. Вони дають змогу оцінити, в якому напрямку рухався акселерометр під час запису, і використовуються для обчислення більш складних метрик, таких як ENMO чи кут нахилу. Змінна `enmo` (Euclidean Norm Minus One) обчислюється як квадратний корінь із суми квадратів прискорення вздовж осей `X`, `Y` і `Z`, з відніманням одиниці та округленням до нуля негативних значень. ENMO є поширеним показником активності, що використовується для виявлення періодів руху або його відсутності. `Angle-Z` (змінна `anglez`) є метрикою, яка обчислює кут нахилу руки відносно горизонтальної площини. Це значення визначається на основі компонентів прискорення й дозволяє аналізувати положення тіла або руки учасника під час спостережень.

Змінна `non-wear_flag` ідентифікує періоди, коли пристрій було знято з руки. Значення 0 означає, що акселерометр носився, а 1 — що пристрій був знятий. Визначення базується на аналізі стандартного відхилення та діапазону

акселерометричних даних відповідно до методології GGIR. Змінна `light` відображає рівень освітленості, виміряний у люксах, що дозволяє досліджувати взаємозв'язок між рівнем фізичної активності та умовами освітлення, а також аналізувати добові цикли активності.

Змінна `battery_voltage` вимірює напругу батареї пристрою в мілівольтах. Вона забезпечує моніторинг технічного стану пристрою та допомагає виявляти періоди, коли дані могли бути спотворені через низький заряд батареї. Змінна `time_of_day` представляє час доби, коли було зібрано дані, у форматі години / хвилини / секунди. Це дозволяє аналізувати добові ритми активності, а також корелювати фізичну активність із часовими інтервалами.

Змінна `weekday` представляє день тижня у вигляді цілого числа: 1 відповідає понеділку, а 7 — неділі. Вона дозволяє вивчати зміни у фізичній активності залежно від дня тижня, наприклад, між робочими днями та вихідними. Змінна `quarter` вказує на квартал року, що дозволяє оцінювати сезонні варіації у фізичній активності, наприклад, залежно від погодних умов або звичних графіків. Змінна `relative_date_PCIAT` показує кількість днів від моменту тестування PCIAT. Негативні значення означають, що дані акселерометрії були зібрані до проведення тесту, а позитивні — після. Це дозволяє вивчати, як часовий проміжок між збором даних та тестом може впливати на результати. Такі багатовимірні дані надають широкий спектр можливостей для аналізу фізичної активності, її зв'язку із поведінковими моделями та прогнозування потенційної залежності від інтернету. Для виконання завдання застосовувався широкий набір бібліотек, що забезпечували обробку даних, побудову моделей, візуалізацію та оптимізацію процесів.

Бібліотека NumPy використовувалася для числових операцій, таких як лінійна алгебра, обробка масивів і статистичні обчислення. Pandas забезпечила маніпуляцію структурованими даними через DataFrame, що включало очищення, фільтрацію, групування та обчислення агрегованих показників. Для роботи з великими наборами даних використовувалася Polars, яка завдяки паралельним обчисленням зменшила час виконання операцій з великими DataFrame.

Для візуалізації даних застосовувалися Matplotlib для побудови базових графіків і Seaborn для складних статистичних візуалізацій, зокрема теплокарт та регресійних графіків.

Побудова моделей машинного навчання здійснювалася за допомогою бібліотек LightGBM, XGBoost і CatBoost, що реалізують алгоритми градієнтного бустингу та забезпечують високу точність у задачах класифікації та регресії. Оптимізацію моделей виконували за допомогою SciPy, зокрема функції minimize, для налаштування порогів і максимізації метрик, таких як квадратичний зважений каппа-коефіцієнт. Scikit-learn забезпечувала крос-валідацію, оцінку моделей і розрахунок метрик продуктивності.

Паралелізація обчислень виконувалася за допомогою ThreadPoolExecutor, а TQDM забезпечувала відображення прогресу виконання у вигляді прогрес-барів, що спрощувало моніторинг тривалих операцій. Для покращення читабельності виведення в консолі застосовувалася бібліотека Colorama. Бібліотека OS забезпечувала взаємодію з файловою системою для маніпулювання шляхами до файлів та доступу до ресурсів, необхідних для обробки даних.

Цей комплекс бібліотек забезпечив ефективну обробку даних, побудову точних моделей і візуалізацію результатів, що дозволило досягти високої якості аналізу та зручності виконання завдання.

3.2. Попередня обробка даних: методи очищення та аналіз

В процесі попередньої обробки файли, що містили дані з акселерометрів, були завантажені і агреговані до щоденних даних, для оптимізації кількості даних в фінальному наборі. Під час аналізу даних учасників було з'ясовано відносну рівномірність даних. Учасники представлені у всіх чотирьох сезонах — зима, весна, літо та осінь, що свідчить про збалансованість вибірки без суттєвого домінування якогось окремого сезону. Таке рівномірне представлення сезонів у

вибірці дозволяє уникнути сезонного зміщення, що може бути корисним для подальшого аналізу, особливо якщо сезон вступу корелює з поведінковими чи іншими характеристиками учасників.

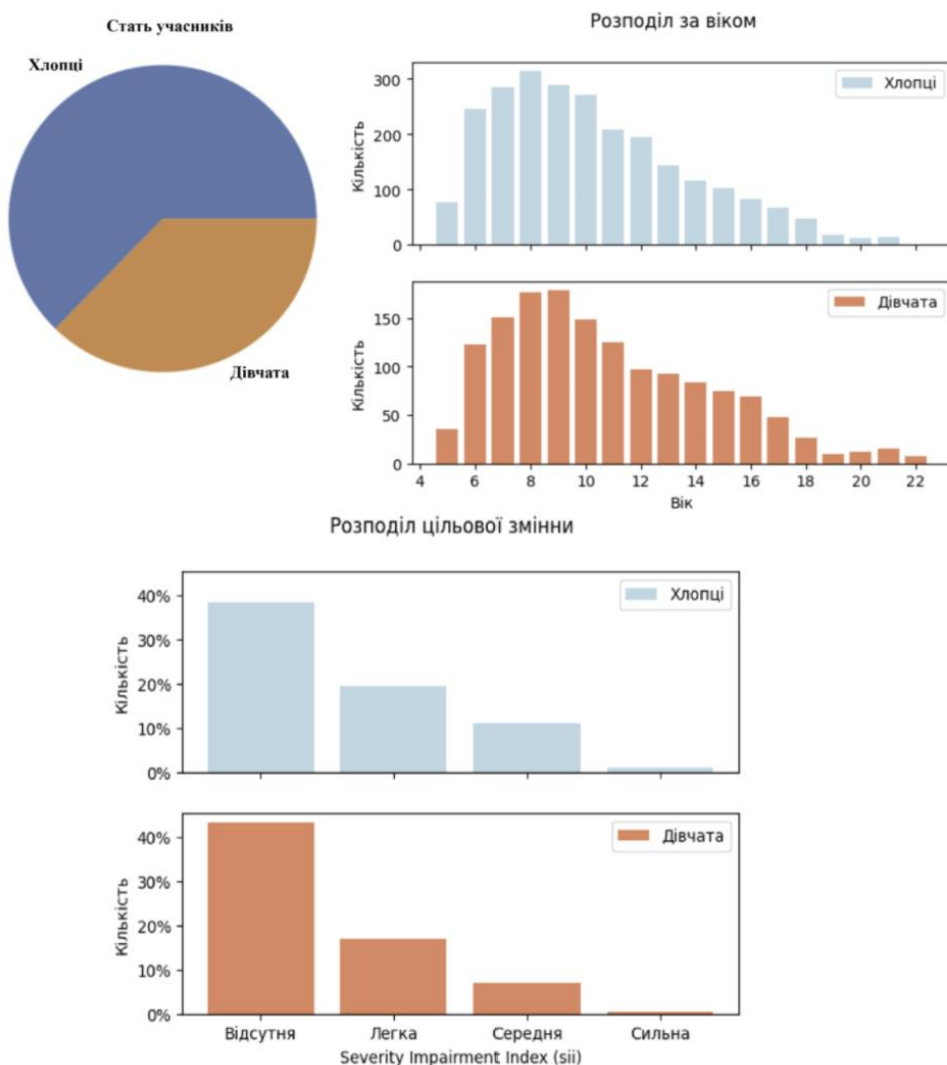


Рисунок 3.3 - Графіки статистичної відмінності між групами за статтю

Джерело: розроблено автором

На основі представлених графіків можна зробити кілька висновків про демографічні характеристики учасників, їхній віковий розподіл та розподіл цільової змінної, яка відображає рівень інтернет залежності (Severity Impairment Index, SII).

Розподіл статей учасників демонструє певний дисбаланс, оскільки більшість вибірки складають хлопці, тоді як дівчата представлені меншою часткою. Такий дисбаланс може впливати на загальні результати аналізу і потребує врахування при

інтерпретації результатів моделей, особливо якщо залежність між інтернет залежністю та статтю виявиться суттєвою.

Віковий розподіл показує суттєву варіативність серед учасників як серед хлопців, так і серед дівчат. Найбільша частка учасників припадає на вікові групи від 6 до 12 років, після чого кількість поступово зменшується з віком. Цей тренд є типовим для досліджень, орієнтованих на молодшу аудиторію, оскільки інтернет залежність частіше досліджується серед дітей і підлітків. При цьому віковий розподіл у хлопців і дівчат має подібну форму, хоча абсолютна кількість хлопців у кожній віковій групі перевищує кількість дівчат.

Розподіл цільової змінної вказує на значну кількість учасників, які не демонструють ознак інтернет залежності. Частка учасників із легким рівнем залежності є помітно меншою, а середній і серйозний рівні залежності представлені значно рідше. Цей розподіл може свідчити про те, що проблема інтернет залежності хоча й поширена, але частіше проявляється у відносно легкій формі. Аналогічна структура розподілу SII спостерігається як серед хлопців, так і серед дівчат, хоча абсолютна кількість хлопців із кожною категорією SII перевищує кількість дівчат.

Для аналізу зв'язків цільової змінної і незалежних змінних було проведено кореляційний аналіз.

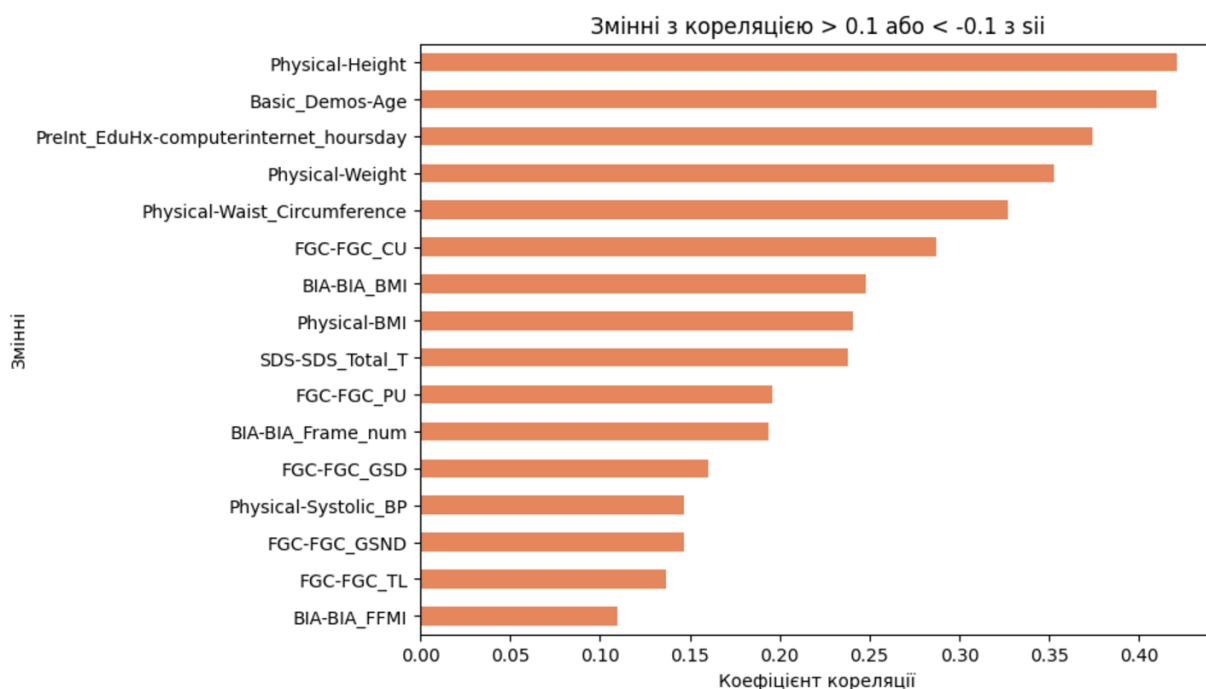


Рисунок 3.4 - Графік величини кореляції змінних з цільовою

Джерело: розроблено автором

Аналіз кореляційних коефіцієнтів показав, що серед представлених змінних найсильніший зв'язок із цільовою змінною SII спостерігається у фізичних та поведінкових показниках, таких як зріст, вік і кількість годин, проведених за комп'ютером чи в інтернеті. Зокрема, змінна Physical-Height демонструє найвищу кореляцію (0.420765), що може бути опосередкованим відображенням вікових особливостей, адже вік також має високу кореляцію (0.409559). Це підтверджує, що віковий фактор є ключовим у визначенні рівня інтернет залежності.

Серед інших змінних суттєву кореляцію продемонстрували фізичні характеристики, такі як вага, об'єм талії та індекс маси тіла. Це вказує на потенційний зв'язок між фізичною активністю, станом здоров'я та інтернет залежністю. Поведінкові фактори, зокрема змінна, що відображає кількість годин за комп'ютером, мають значний вплив, адже частота використання технологій прямо корелює з рівнем залежності.

Загалом, результати свідчать про те, що інтернет залежність є багатовимірним явищем, яке залежить від вікових, фізичних та поведінкових характеристик. Найвища кореляція зі зростом і віком підкреслює важливість

урахування демографічних особливостей у прогнозуванні інтернет залежності, тоді як фізичні та поведінкові показники доповнюють загальну картину.

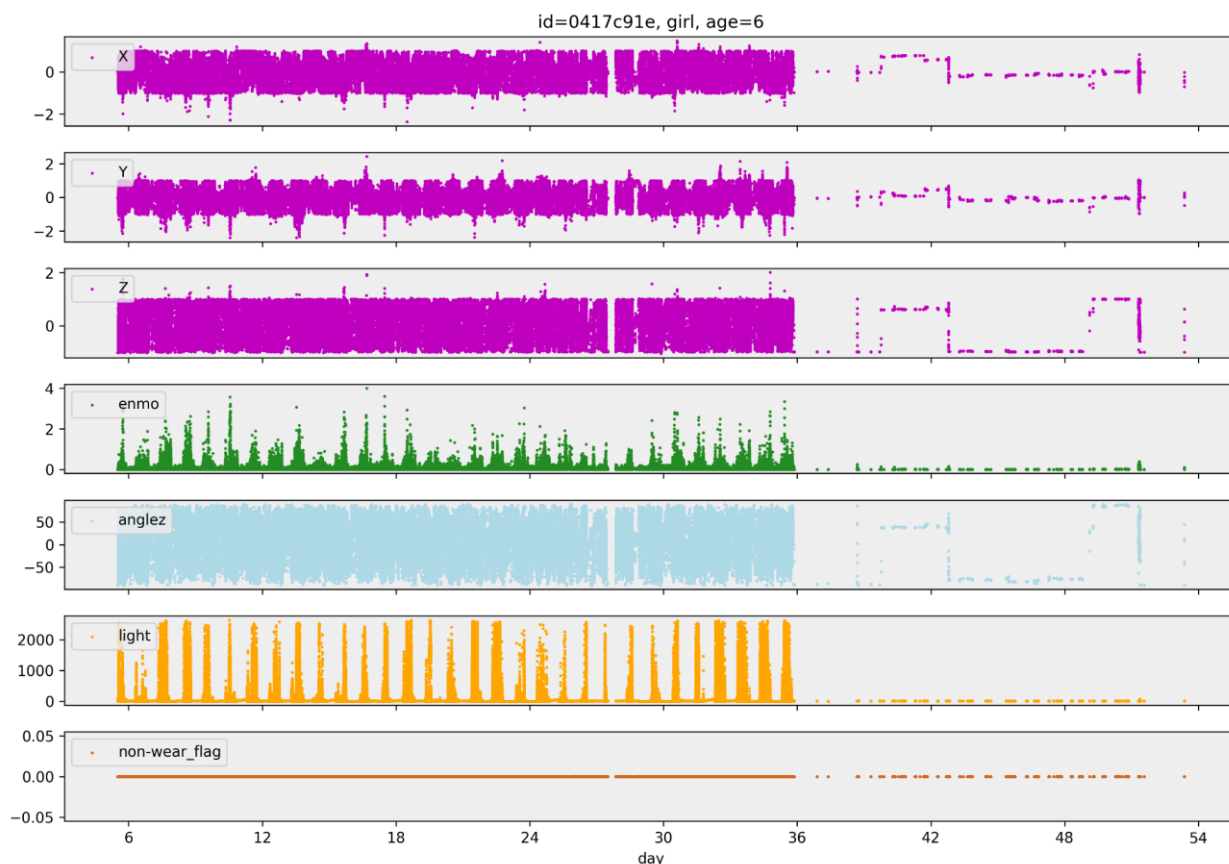


Рисунок 3.5 - Візуалізація даних з акселерометрів

Джерело: розроблено автором

Дані з акселерометрів було агреговано до щоденних показників для оптимізації розрахунків і поєднано з інформацією основної таблиці про учасників.

Рисунок 3.5 ілюструє дані з акселерометрів, зібрані для учасника дослідження який є дівчинкою віком 6 років. Графік складається з декількох панелей, кожна з яких представляє окремий параметр, отриманий з акселерометра. Дані відображені по днях, охоплюючи період приблизно в 50 днів.

На перших трьох панелях показані компоненти прискорення по осях X, Y і Z. Ці дані відображають рухи учасника у просторі, причому зміни в амплітуді сигналу вказують на різну інтенсивність активності. Такі параметри дають змогу оцінювати фізичну активність у різні періоди часу.

Четверта панель демонструє метрику ENMO, яка є інтегральним показником рухової активності. Тут видно регулярні піки, які відповідають періодам

активності, наприклад, іграм чи прогулянкам, а також значні періоди зниженого руху, що можуть вказувати на сон або відпочинок.

П'ята панель відображає показник Angle-Z, який характеризує положення акселерометра відносно горизонтальної площини. Ця змінна допомагає визначити, в якій позиції перебуває рука учасника, наприклад, коли вона піднята чи опущена.

Шоста панель представляє рівень освітленості (Light), виміряний у люксах. Тут чітко простежується денна сезонність: рівень світла значно підвищується в денні години і знижується вночі, що відображає природні цикли освітлення. На останній панелі зображено прапор non-wear_flag, який вказує, коли акселерометр не носили. Нульові значення означають, що пристрій був на учаснику, тоді як ненульові — періоди, коли пристрій було знято. Дані свідчать про те, що акселерометр носили переважно регулярно, але є кілька періодів, коли пристрій був знятий.

У загальному плані дані демонструють типові добові цикли активності та освітленості. Наприклад, у денний час спостерігається підвищена рухова активність та високий рівень світла, тоді як уночі активність значно знижується, а освітлення мінімальне. Така структура даних дозволяє робити висновки про ритм життя, звички та рівень фізичної активності учасника.

3.3. Тренування моделей: алгоритми, параметри та оцінка ефективності

У рамках підготовки даних для дослідження прогнозування схильності до інтернет залежності було реалізовано комплексний підхід до обробки, аналізу та трансформації вихідного набору даних. Основною метою цих етапів було створення високоякісного та інформативного набору ознак, який дозволяє побудувати точні та узагальнені моделі прогнозування.

Одним із ключових джерел даних стали акселерометричні тимчасові ряди, які відображають фізичну активність учасників протягом кількох днів. Дані зберігалися у форматі Parquet і були організовані у вигляді файлів, де кожен файл відповідав одному учаснику. Для ефективної роботи з такими даними було реалізовано автоматизований процес обчислення базових статистичних характеристик для кожного тимчасового ряду, зокрема середнього значення, стандартного відхилення та інших метрик. Це дозволило конденсувати багатовимірні ряди до компактної форми, що значно зменшило обсяг даних і спростило їх інтеграцію з іншими джерелами.

Для подальшого зменшення розмірності даних акселерометрів і виділення їх найбільш значущих характеристик було застосовано нейронну мережу автоенкодера. Через процес кодування-декодування автоенкодер мінімізував похибку реконструкції, виділяючи латентні змінні, які додавалися до основного набору даних як нові ознаки. Це забезпечило компактність і збереження найбільш важливої інформації про рухову активність, одночасно оптимізуючи обчислювальну ефективність.

Додатковим етапом стала генерація нових ознак шляхом взаємодії між існуючими характеристиками. Серед створених показників були індекс маси тіла, скорегований на вік учасника, кількість годин, проведених в інтернеті, помножена на вік, співвідношення м'язової маси до жирової тканини, рівень гідратації, а також інші інтерактивні метрики. Ці ознаки відображали складні взаємозв'язки між фізичними, демографічними та поведінковими характеристиками учасників і були спрямовані на підвищення прогностної спроможності моделі.

BMI_Internet_Hours	BFP_BMI	FFMI_BFP	FMI_BFP	LST_TBW	BFP_BMR	BFP_DEE	BMR_Weight	DEE_Weight	SMM_Height	Muscle_to_Fat	Hydration_Status
50.631947	0.545865	1.499679	0.332267	1.190475	8591.822097	13746.944840	18.356260	29.370079	0.424811	6.383063	0.643522
0.000000	0.282883	3.229888	0.305154	1.458119	3719.320478	5950.914352	20.362087	32.579348	0.321056	12.718037	0.588157
33.297392	0.923872	0.840729	0.241670	1.231775	19528.656877	33346.609152	14.634960	24.990265	0.488767	6.473938	0.608640
0.000000	1.028971	0.747651	0.224196	1.281264	21298.377749	36207.411592	13.865564	23.571569	0.472854	6.274343	0.563684
69.455460	2.174098	0.282649	0.177312	1.328209	79057.011034	121546.184578	11.268078	17.324103	0.760160	4.731007	0.525843
0.000000	2.251718	0.245510	0.198595	1.262516	90468.027355	135701.701175	11.862478	17.793672	0.594629	2.621003	0.562625

Рисунок 3.6 - Нові утворені змінні

Джерело: розроблено автором

Пропущені значення в даних були імітовані за допомогою алгоритму KNNImputer, який використовував інформацію про найближчих сусідів для заповнення відсутніх даних. Це дозволило зберегти контекст даних і зменшити втрати інформації. Записи з великою кількістю пропусків, які неможливо було коректно заповнити, видалялися, щоб забезпечити якість фінального набору. Для узгодженості числові змінні були стандартизовані за допомогою StandartScaler, що особливо важливо для алгоритмів машинного навчання, чутливих до масштабу ознак.

Після завершення обробки даних усі таблиці були об'єднані в один набір, з 124 змінних який включав статистичні характеристики акселерометрів, латентні змінні автоенкодера, основні демографічні, фізичні та поведінкові характеристики, а також нові інтерактивні ознаки. Малозначущі змінні, такі як сезонні характеристики, були вилючені, щоб уникнути надлишковості. Це створило цілісний, високоякісний набір даних, готовий для моделювання. Такий підхід дозволив забезпечити високу якість даних і підвищити точність моделі, що є критично важливим для досягнення мети дослідження.

У процесі навчання та оцінювання моделей було застосовано сучасні підходи машинного навчання, зокрема використання градієнтного бустингу та ансамблевих методів, які дозволяють досягти високої точності прогнозів. Основна мета — побудова моделі, яка здатна передбачити показник схильності до інтернет залежності, зберігаючи при цьому узгодженість результатів з особливостями цільової змінної, яка має порядкову природу.

Для навчання було обрано кілька моделей градієнтного бустингу: LightGBM, XGBoost і CatBoost. Кожна з цих моделей має свої переваги: LightGBM забезпечує високу швидкість навчання на великих наборах даних, XGBoost є надзвичайно потужним для структурованих даних, а CatBoost оптимізований для роботи з категоріальними змінними, що робить його особливо корисним у контексті дослідження.

У процесі моделювання було обрано специфічні параметри для моделей LightGBM, XGBoost та CatBoost, що оптимізували їхню продуктивність у задачі

прогнозування. Параметри були налаштовані з урахуванням структури даних та завдання, щоб забезпечити баланс між точністю, швидкістю обчислень та здатністю моделей до узагальнення, за допомогою GridSearchCV .

```
Params = {
    'learning_rate': 0.046,
    'max_depth': 12,
    'num_leaves': 478,
    'min_data_in_leaf': 13,
    'feature_fraction': 0.893,
    'bagging_fraction': 0.784,
    'bagging_freq': 4,
    'lambda_l1': 10,
    'lambda_l2': 0.01,
    'device': 'gpu'
}

CatBoost_Params = {
    'learning_rate': 0.05,
    'depth': 6,
    'iterations': 200,
    'random_seed': SEED,
    'verbose': 0,
    'l2_leaf_reg': 10,
    'task_type': 'GPU'
}

XGB_Params = {
    'learning_rate': 0.05,
    'max_depth': 6,
    'n_estimators': 200,
    'subsample': 0.8,
    'colsample_bytree': 0.8,
    'reg_alpha': 1,
    'reg_lambda': 5,
    'random_state': SEED,
    'tree_method': 'gpu_hist',
}
```

Рисунок 3.7 – Використані параметри моделі

Джерело: розроблено автором

Ансамблювання через Voting Regression додатково підвищило стабільність та узгодженість результатів. Параметри моделей були налаштовані з урахуванням загальних вимог до задачі, таких як стабільність, здатність до узагальнення та швидкість обчислень.

Для LightGBM, XGBoost та CatBoost низьке значення параметра `learning_rate` (0.046 для LightGBM, 0.05 для XGBoost та CatBoost) забезпечує поступове коригування ваг моделі, що зменшує ризик перенавчання та сприяє стабільному навчальному процесу.

Параметр `max_depth` обмежує глибину дерев (12 для LightGBM та 6 для XGBoost і CatBoost), що запобігає надмірній складності моделей і покращує їхню здатність до узагальнення. Для LightGBM параметри `num_leaves: 478` та `min_data_in_leaf: 13` дозволяють ефективно захоплювати складні закономірності, забезпечуючи баланс між складністю моделі та уникненням створення надмірно специфічних вузлів.

Регуляризаційні параметри, такі як `lambda_l1`, `lambda_l2` у LightGBM (10 та 0.01 відповідно) та `reg_alpha`, `reg_lambda` у XGBoost (1 та 5 відповідно), були налаштовані для контролю над вагами ознак, що сприяє підвищенню стійкості

моделі та зменшенню ймовірності перенавчання. CatBoost використовує аналогічний параметр `l2_leaf_reg`: 10 для зменшення складності дерев.

Метрики обчислювальної ефективності враховувалися за допомогою використання часткових підвбірок ознак і даних у LightGBM та XGBoost (параметри `feature_fraction`: 0.893, `bagging_fraction`: 0.784, `bagging_freq`: 4 для LightGBM і `subsample`: 0.8, `colsample_bytree`: 0.8 для XGBoost). Це знижувало кореляцію між деревами та запобігало перенавчанню.

GPU-оптимізовані алгоритми використовувалися для всіх трьох моделей, параметри `device`: 'gpu' у LightGBM, `tree_method`: 'gpu_hist' у XGBoost, `task_type`: 'GPU' у CatBoost, що суттєво пришвидшило процес навчання, особливо при роботі з великими обсягами даних.

Для CatBoost параметр `iterations`: 200 забезпечує ефективне навчання без ризику перенавчання, а його здатність обробляти категоріальні ознаки без попередньої обробки забезпечує точність при мінімальних витратах на підготовку даних.

Всі параметри були обрані з метою досягнення оптимального балансу між продуктивністю, стійкістю та здатністю до узагальнення, що є критично важливим для складних задач прогнозування.

З метою підвищення точності було створено ансамблевую модель на основі Voting Regressor, яка об'єднує прогнози зазначених моделей. Цей підхід дозволив використати сильні сторони кожної моделі, зменшивши при цьому ризик переобчислення або залежності від однієї специфічної моделі. Ансамблева модель забезпечує більш стабільні прогнози, оскільки комбінує результати кількох джерел.

Для розподілу даних між тренувальними та валідаційними вибірками використовувався стратифікований метод перехресної перевірки K-Folds, що гарантує баланс класів у кожній підмножині. Це дозволило оцінити узагальнюючу здатність моделі, зменшивши ризик переобчислення та надаючи змогу більш точно оцінити її ефективність.

Оцінювання моделей здійснювалося за допомогою метрики Quadratic Weighted Kappa, яка враховує порядковий характер цільової змінної. Ця метрика

дозволяє оцінити, наскільки добре модель передбачає схильність до інтернет залежності, беручи до уваги як точність, так і значення помилки для кожної категорії.

Для підвищення точності було виконано оптимізацію порогів, які використовуються для конвертації безперервних значень прогнозів у дискретні категорії. Цей процес базувався на мінімізації функції втрат за допомогою алгоритму Nelder-Mead. Оптимізація дозволила врахувати специфіку розподілу даних і краще адаптувати модель до вимог метрики QWK.

Отже, використання поєднання окремих моделей градієнтного бустингу, ансамблевих методів, стратифікованої перехресної перевірки та оптимізації порогів забезпечило високу точність та узгодженість моделі з даними. Цей підхід дозволяє не лише отримати надійні прогнози, а й забезпечує гнучкість у роботі з різними типами змінних і структурою даних.

```

Training Folds: 100%|██████████| 5/5 [00:53<00:00, 10.74s/it]
Mean Train QWK --> 0.7595
Mean Validation QWK ---> 0.3926

----> || Optimized QWK SCORE :: 0.457

[22]:
      id  sii
0  00008ff9  1
1  000fd460  0
2  00105258  0
3  00115b9f  0
4  0016bb22  0
5  001f3379  1
6  0038ba98  0
7  0068a485  0
8  0069fbcd  1
9  0083e397  0
10 0087dd65  0
11 00abe655  0
12 00aa59c9  1

```

Рисунок 3.8 - Оцінка результатів прогнозування

Джерело: розроблено автором

Отримані результати показують, що середній QWK для тренувальних даних становить 0.7595, що свідчить про гарну узгодженість моделі з тренувальним набором. Проте, середній QWK для валідаційних даних значно нижчий і дорівнює

0.3926, що може вказувати на проблему узагальнення моделі або її чутливість до специфіки валідаційних даних. Після оптимізації порогів показник QWK зріс до 0.457, що свідчить про певне покращення прогностної здатності моделі.

Отримані результати свідчать про потенціал розробленої моделі для застосування у задачах моніторингу стану здоров'я та оцінки ризиків інтернет залежності. Висока узгодженість моделі з тренувальними даними свідчить про здатність алгоритму ефективно виявляти закономірності в структурі даних, а застосування оптимізації порогів дозволило частково вирішити проблему узагальнення, покращивши QWK на валідаційних даних до 0.457. Це підтверджує доцільність використання моделі для задач, де важливо забезпечити точність прогнозів при обмежених даних та їх нерівномірному розподілі.

Дану модель можна використовувати для моніторингу стану здоров'я у закладах освіти, центрах громадського здоров'я, або в умовах, де існує обмежений доступ до професійних медичних працівників. Завдяки можливості автоматизованої оцінки ризиків, модель здатна зменшити навантаження на медичні установи, надаючи попередню діагностику та ідентифікуючи групи ризику для подальшого цільового втручання. Це особливо актуально для регіонів із браком людських ресурсів або великим навантаженням на системи охорони здоров'я.

Крім того, модель може бути інтегрована в системи електронного здоров'я та мобільні додатки для моніторингу психічного здоров'я. Завдяки ефективності моделі в обробці даних різного формату, її можна використовувати у масштабованих системах для моніторингу стану пацієнтів у реальному часі. Це дозволяє своєчасно виявляти потенційні проблеми та здійснювати запобіжні заходи до виникнення серйозних наслідків. Застосування моделі також може бути корисним у проведенні великих популяційних досліджень для аналізу поширеності інтернет залежності серед різних груп населення. Це надає можливість виявляти тенденції у здоров'ї населення, формувати рекомендації для політики у сфері охорони здоров'я, а також сприяти підвищенню обізнаності щодо проблеми інтернет залежності серед громадськості.

Загалом, розроблена модель є важливим інструментом для автоматизації моніторингу стану здоров'я та профілактики інтернет залежності, що дозволяє покращити ефективність медичних та соціальних систем, сприяючи підвищенню якості життя населення.

ВИСНОВКИ

Актуальність дослідження зумовлена стрімким поширенням інтернет залежності, що впливає на психічне здоров'я, соціальну адаптацію та продуктивність населення, особливо серед молоді. Сучасний цифровий світ вимагає розробки ефективних методів виявлення схильності до цього стану, що дозволить зменшити його негативний вплив. Використання інноваційних підходів на основі машинного навчання забезпечує можливість прогнозування ризиків на ранніх етапах, сприяючи своєчасному втручанню та профілактиці, що є особливо важливим для формування здорового суспільства.

У межах дослідження було здійснено комплексний аналіз даних, їх передобробку, створення нових ознак, побудову моделей машинного навчання та оцінювання їхньої ефективності з метою прогнозування схильності до інтернет залежності. Використання сучасних методів, таких як градієнтний бустинг (LightGBM, XGBoost, CatBoost), ансамблеві підходи (Voting Regressor) та оптимізація порогів, дозволило створити узгоджену систему, яка враховує багатофакторну природу досліджуваного явища.

Результати моделювання показали, що середній показник Quadratic Weighted Карра на тренувальних даних становив 0.7595, що свідчить про високу точність моделей у навчанні на основному наборі. Такий результат вказує на здатність моделей ефективно виявляти патерни в даних і адекватно враховувати взаємозв'язки між фізичними, демографічними та поведінковими ознаками. Однак середній показник QWK на валідаційних даних був значно нижчим — 0.3926. Це може бути пов'язано з певними обмеженнями у якості даних, розподілом цільової

змінної або з тим, що моделі, попри високу продуктивність на тренувальній вибірці, демонструють обмежену здатність узагальнювати знання на нові дані.

Застосування процедури оптимізації порогів дозволило підвищити кінцевий показник QWK до 0.457. Це свідчить про те, що адаптація порогових значень для класифікації на основі розподілу даних допомагає враховувати специфіку порядкової змінної і покращує узгодженість між передбачуваними та реальними значеннями. Проте, отриманий результат також вказує на те, що модель усе ще має обмеження у точності прогнозування, особливо на складних або нерівномірних вибірках.

Одним із ключових досягнень дослідження стало створення інтегрованого набору даних, який поєднує різноманітні фізичні, демографічні та поведінкові характеристики учасників, а також кодовані представлення даних акселерометрів. Генерація нових ознак на основі взаємодії змінних значно розширила інформативність моделі, дозволяючи враховувати приховані взаємозв'язки між характеристиками учасників. Крім того, імітація пропущених значень за допомогою KNNImputer дозволила зберегти цілісність набору даних, що стало важливим фактором для підвищення ефективності моделей.

Результати дослідження вказують на потенціал запропонованого підходу для прогнозування схильності до інтернет залежності, однак підкреслюють необхідність подальшої роботи над удосконаленням моделей. Зокрема, перспективним напрямом є використання більш складних ансамблевих методів, обробка дисбалансу у вибірках та дослідження додаткових змінних, які можуть впливати на результати. Загалом, проведене дослідження демонструє, що застосування методів машинного навчання є ефективним інструментом для вирішення складних соціальних проблем, таких як прогнозування інтернет залежності, та формує основу для подальших наукових розробок у цій галузі.

Незважаючи на досягнуті результати, проведене дослідження має певні обмеження, які впливають на загальну ефективність і узагальнюваність отриманих моделей. Одним із ключових обмежень є нерівномірність розподілу даних у вибірці, зокрема для цільової змінної (s_{ii}), яка має порядкову природу.

Незбалансованість класів у цільовій змінній обмежує здатність моделі адекватно прогнозувати менш поширені категорії, такі як "Severe", що, своєю чергою, знижує загальну продуктивність моделі. Майбутні дослідження можуть зосередитися на розв'язанні цієї проблеми шляхом застосування методів збалансування даних, таких як техніки надсемплінгу або генеративні підходи для синтетичного розширення вибірки.

Ще одним значущим обмеженням є значна кількість пропущених значень у даних. Хоча використання KNNImputer дозволило зберегти основну інформацію, такий підхід може бути недостатньо точним у випадках, коли відсутність даних не є випадковою. Це може вплинути на якість моделі, особливо якщо пропущені значення корелюють із цільовою змінною. Майбутні дослідження можуть спрямувати зусилля на розробку більш адаптивних методів заповнення пропусків, таких як моделювання пропущених значень на основі нейронних мереж або більш складних статистичних підходів.

Крім того, акселерометричні дані, які становлять важливу частину набору, були представлені у вигляді узагальнених статистичних характеристик і латентних змінних, отриманих за допомогою автоенкодера. Цей підхід дозволяє зменшити розмірність, однак може втратити частину інформації про динамічні патерни рухової активності. Подальші дослідження можуть зосередитися на використанні більш складних моделей для аналізу часових рядів, таких як рекурентні нейронні мережі (RNN) або трансформери, що дадуть змогу зберегти послідовні аспекти даних і потенційно покращити точність прогнозів.

Ще однією перспективою є дослідження додаткових факторів, які можуть впливати на схильність до інтернет залежності. Наразі аналіз зосереджено на фізичних, демографічних та поведінкових характеристиках, однак включення даних про соціальне середовище, психологічний стан або взаємодію з технологіями може значно розширити можливості моделі. Наприклад, інтеграція даних із соціальних мереж, поведінкових анкет або сенсорів інших типів може надати більш повну картину.

Також важливим напрямом є вдосконалення метрик оцінювання. Використання Quadratic Weighted Кappa забезпечує врахування порядкової природи цільової змінної, однак більш детальний аналіз помилок моделі може включати застосування інших метрик, таких як F1-міра для кожного класу, або аналіз розподілу прогнозованих і реальних значень для виявлення вузьких місць моделі.

Загалом, обмеження цього дослідження окреслюють низку перспективних напрямів для подальшої роботи, зокрема вдосконалення методів обробки даних, інтеграція нових джерел інформації, використання більш потужних алгоритмів для аналізу часових рядів та розробка адаптивних підходів для роботи з дисбалансами у вибірці. Ці заходи можуть суттєво підвищити як якість прогнозування, так і практичну значущість моделей для вирішення проблеми інтернет залежності серед молоді.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Senormancı O., Konkan R., Sungur M. Internet Addiction and Its Cognitive Behavioral Therapy. Standard and Innovative Strategies in Cognitive Behavior Therapy. Anatolian Journal of Psychiatry. 2012. Vol. 11, no. 3. URL: <http://dx.doi.org/10.5772/26265> (date of access: 19.11.2024).
2. Zhijun Guo, Laigang Zhang. A new model of Internet addiction tendency based on machine learning// Third International Conference on Computer Science and Communication Technology (ICCSCT 2022). 2022. Proceedings Volume 12506. №125061X. URL: <https://doi.org/10.1117/12.2662545> (date of access: 28.12.2022).
3. Hua, Yuming & Guo, Junhai & Zhao, Hua. (2015). Deep Belief Networks and deep learning. Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things, ICIT 2015. 1-4. 10.1109/ICAIOT.2015.7111524. URL: https://www.researchgate.net/publication/282318042_Deep_Belief_Networks_and_deep_learning (date of access: 28.12.2022).
4. Shuang Zhang, Huisi Yu. Online addiction analysis and identification of students by applying gd-LSTM algorithm to educational behaviour data // Journal of Intelligent Systems. 2024. T. 33. №1. C. 20230102. URL: <https://doi.org/10.1515/jisys-2023-0102> (date of access: 28.12.2022).
5. W. Zhang, H. Quan, W. Zhang and Z. Li, "Short-Term Wind Power Interval Prediction Based on GD-LSTM and Bootstrap Techniques," 2022 IEEE 5th International Electrical and Energy Conference (CIEEC), Nangjing, China, 2022, pp. 2626-2631. URL: <https://ieeexplore.ieee.org/document/9846316> (date of access: 28.12.2022).
6. Wright G. What is ARPANET and what's its significance?. Search Networking. URL: <https://www.techtarget.com/searchnetworking/definition/ARPANET#:~:text=The>

[%20U.S.%20Advanced%20Research%20Projects,and%20finally%20decommissi
oned%20in%201989](#). (date of access: 20.11.2024).

7. Young, K. S. (1998). Internet addiction: The emergence of a new clinical disorder. *CyberPsychology & Behavior*, 1(3), 237–244. <https://doi.org/10.1089/cpb.1998.1.237> (date of access: 20.11.2024)
8. Bell CC. DSM-IV: Diagnostic and Statistical Manual of Mental Disorders. *JAMA*. 1994;272(10):828–829. doi:10.1001/jama.1994.03520100096046. URL: <https://jamanetwork.com/journals/jama/article-abstract/379036> (date of access: 20.11.2024).
9. Mittal, Aditi & Bansal, Nitin. (2019). The Impact of Technological Advancement on Digital Addiction Level of Youth in India. *International Journal of Innovative Technology and Exploring Engineering*. 8. 775-783. 10.35940/ijitee.L1419.10812S19. URL:<http://dx.doi.org/10.35940/ijitee.L1419.10812S19> (date of access: 20.11.2024).
10. Anderson, Keith. (2001). Internet Use Among College Students: An Exploratory Study. *Journal of American college health: J of ACH*. 50. 21-6. 10.1080/07448480109595707. URL:<http://dx.doi.org/10.1080/07448480109595707> (date of access: 20.11.2024).
11. Griffiths, Mark. (1999). Internet Addiction: Fact or Fiction?. *The Psychologist*. 12. 246-250. URL: [https://www.researchgate.net/publication/232557569_Internet_Addiction_Fact_o
r_Fiction](https://www.researchgate.net/publication/232557569_Internet_Addiction_Fact_or_Fiction) (date of access: 21.11.2024).
12. Public health implications of excessive use of the Internet and other communication and gaming platforms. World Health Organization (WHO). URL: [https://www.who.int/news/item/13-09-2018-public-health-implications-of-
excessive-use-of-the-internet-and-other-communication-and-gaming-platforms](https://www.who.int/news/item/13-09-2018-public-health-implications-of-excessive-use-of-the-internet-and-other-communication-and-gaming-platforms) (date of access: 21.11.2024).
13. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing; 2013. URL:

- <https://psycnet.apa.org/doi/10.1176/appi.books.9780890425596> (date of access: 21.11.2024).
- 14.Liu, L., Yao, Y. W., Li, C. R., Zhang, J. T., Xia, C. C., Lan, J., Ma, S. S., Zhou, N., & Fang, X. Y. (2018). The Comorbidity Between Internet Gaming Disorder and Depression: Interrelationship and Neural Mechanisms. *Frontiers in psychiatry*, 9, 154. <https://doi.org/10.3389/fpsyt.2018.00154>
 - 15.Laconi, S., Rodgers, R. F., & Chabrol, H. (2014). The measurement of Internet addiction: A critical review of existing scales and their psychometric properties. *Computers in Human Behavior*, 41, 190–202. <https://doi.org/10.1016/j.chb.2014.09.026> (date of access: 21.11.2024).
 - 16.Hussain, Z., & Pontes, H. M. (2018). Personality, internet addiction, and other technological addictions: A psychological examination of personality traits and technological addictions. In B. Bozoglan (Ed.), *Psychological, social, and cultural aspects of internet addiction* (pp. 45–71). Information Science Reference/IGI Global. <https://doi.org/10.4018/978-1-5225-3477-8.ch003> (date of access: 21.11.2024).
 - 17.Chemnad, K., Aziz, M., Abdelmoneium, A. et al. Adolescents' Internet addiction: Does it all begin with their environment?. *Child Adolesc Psychiatry Ment Health* 17, 87 (2023). <https://doi.org/10.1186/s13034-023-00626-7> (date of access: 21.11.2024).
 - 18.Hoffman BR, Unger JB. The Role of Culture in Addiction. In: Sussman S, ed. *The Cambridge Handbook of Substance and Behavioral Addictions*. Cambridge Handbooks in Psychology. Cambridge University Press; 2020:171-181. URL: <https://doi.org/10.1017/9781108632591.018> (date of access: 21.11.2024).
 - 19.Sun, S., Yang, J., Chen, Y., Miao, J., & Sawan, M. (2021). EEG Signals Based Internet Addiction Diagnosis Using Convolutional Neural Networks. *Applied Sciences*, 12(13), 6297. <https://doi.org/10.3390/app12136297> (date of access: 21.11.2024).

20. Facts and Figures 2022 - Youth Internet use. ITU: Committed to connecting the world. URL: <https://www.itu.int/itu-d/reports/statistics/2022/11/24/ff22-youth-internet-use/> (date of access: 22.11.2024).
21. Islam, M.R. Unplugging the threat: how internet addiction among adolescents undermines learning behavior. *SN Soc Sci* **4**, 161 (2024). <https://doi.org/10.1007/s43545-024-00966-x> (date of access: 22.11.2024).
22. AlHeneidi, H.H., Smith, A.P. (2021). Exploring the Influence of Information Overload, Internet Addiction, and Social Network Addiction, on Students' Well-Being and Academic Outcomes. In: Longo, L., Leva, M.C. (eds) *Human Mental Workload: Models and Applications. H-WORKLOAD 2021. Communications in Computer and Information Science*, vol 1493. Springer, Cham. https://doi.org/10.1007/978-3-030-91408-0_8 (date of access: 22.11.2024).
23. Dong, W., Tang, H., Wu, S. et al. The effect of social anxiety on teenagers' internet addiction: the mediating role of loneliness and coping styles. *BMC Psychiatry* **24**, 395 (2024). <https://doi.org/10.1186/s12888-024-05854-5> (date of access: 22.11.2024).
24. Cheol-Soon Lee, Kwame McKenzie. Socioeconomic and Geographic Inequalities of Internet Addiction in Korean Adolescents. *Psychiatry Investigation*. URL: <https://doi.org/10.4306/pi.2015.12.4.559> (date of access: 22.11.2024).
25. Zhu, Y., Chen, H., Li, J. et al. Effects of different interventions on internet addiction: a systematic review and network meta-analysis. *BMC Psychiatry* **23**, 921 (2023). <https://doi.org/10.1186/s12888-023-05400-9> (date of access: 22.11.2024).
26. Allcott, Hunt, Gentzkow, Matthew, Song, Lena. Digital Addiction. *American Economic Review* **112**, 7, 2424–63, 2022. <https://www.aeaweb.org/articles?id=10.1257/aer.20210867> (date of access: 22.11.2024).
27. Bhargava, V.R. and Velasquez, M. (2021) 'Ethics of the Attention Economy: The Problem of Social Media Addiction', *Business Ethics Quarterly*, **31**(3), pp. 321–359. URL: <https://doi.org/10.1017/beq.2020.32> (date of access: 22.11.2024).

28. Lee, Hae Kook, and Sulki Chung, 'Conceptualization of Internet Addiction Based on the Public Health Perspective', in Marc N. Potenza, Kyle A. Faust, and David Faust (eds), *The Oxford Handbook of Digital Technologies and Mental Health* (2020; online edn, Oxford Academic, 8 Oct. 2020), <https://doi.org/10.1093/oxfordhb/9780190218058.013.8> (date of access: 22.11.2024).
29. Jiang, Q. (2022, September 15). Development and Effects of Internet Addiction in China. *Oxford Research Encyclopedia of Communication*. Retrieved 22 Nov. 2024, from <https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-1142> (date of access: 22.11.2024).
30. Cho, C.-M. (2016) 'South Korea's Efforts to Prevent Internet Addiction', in M. Israelashvili and J.L. Romano (eds.) *The Cambridge Handbook of International Prevention Science*. Cambridge: Cambridge University Press (Cambridge Handbooks in Psychology), pp. 551–573. URL: <https://doi.org/10.1017/9781316104453.024> (date of access: 22.11.2024).
31. Expected Outcomes: Comparative Analysis of Drug Rehabilitation Programs - fundsforNGOs - Grants and Resources for Sustainability. URL: <https://www.fundsforngos.org/proposals/expected-outcomes-comparative-analysis-of-drug-rehabilitation-programs/> (date of access: 22.11.2024).
32. ALL DIGITAL. Enhancing Digital Skills Across Europe. URL: <https://all-digital.org/wp-content/uploads/2023/12/20.12.2023-Digital-Well-Being-Report.pdf> (date of access: 22.11.2024).
33. Addictive behaviours: Gaming disorder. World Health Organization (WHO). URL: <https://www.who.int/news-room/questions-and-answers/item/addictive-behaviours-gaming-disorder> (date of access: 22.11.2024).
34. Bickham, D.S. Current Research and Viewpoints on Internet Addiction in Adolescents. *Curr Pediatr Rep* **9**, 1–10 (2021). <https://doi.org/10.1007/s40124-020-00236-3> (date of access: 22.11.2024).

35. Diotaiuti, P., Mancone, S., Corrado, S., De Risio, A., Cavicchiolo, E., Girelli, L., & Chirico, A. (2022). Internet addiction in young adults: The role of impulsivity and codependency. *Frontiers in Psychiatry*, 13, 893861. <https://doi.org/10.3389/fpsy.2022.893861> (date of access: 22.11.2024).
36. Teichgraeber, H., & Brandt, A. R. (2022). Time-series aggregation for the optimization of energy systems: goals, challenges, approaches, and opportunities. *Optimization Online*. Retrieved from <https://optimization-online.org/wp-content/uploads/2022/01/8753.pdf> (date of access: 23.11.2024).
37. Rostami-Tabar, B., Babai, M. Z., & Syntetos, A. (2021). To aggregate or not to aggregate: Forecasting of finite autocorrelated demand. *arXiv*. Retrieved from <https://arxiv.org/abs/2103.16310> (date of access: 23.11.2024).
38. Rawat, T., & Khemchandani, V. (2017). Feature Engineering (FE) Tools and Techniques for Better Classification Performance. *International Journal of Innovations in Engineering and Technology*, 8(2), 169–176. <http://dx.doi.org/10.21172/ijiet.82.024> (date of access: 23.11.2024).
39. Daniele Micci-Barreca. 2001. A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. *SIGKDD Explor. Newsl.* 3, 1 (July 2001), 27–32. <https://doi.org/10.1145/507533.507538> (date of access: 23.11.2024).
40. Azuaje F. (2006). Review of "Data Mining: Practical Machine Learning Tools and Techniques" by Witten and Frank. *BioMedical Engineering OnLine*, 5, 51. <https://doi.org/10.1186/1475-925X-5-51> (date of access: 23.11.2024).
41. Maaten, Laurens van der, Eric O. Postma and Jaap van den Herik. "Dimensionality Reduction: A Comparative Review." (2008). URL: https://lvdmaaten.github.io/publications/papers/TR_Dimensionality_Reduction_Review_2009.pdf (date of access: 23.11.2024).
42. Sorzano, Carlos & Vargas, Javier & Montano, A.. (2014). A survey of dimensionality reduction techniques. URL: <https://doi.org/10.48550/arXiv.1403.2877> (date of access: 23.11.2024).

43. Birla D. Autoencoders. *Medium*. URL: <https://medium.com/@birla.deepak26/autoencoders-76bb49ae6a8f> (date of access: 29.11.2024).
44. J. Zhai, S. Zhang, J. Chen and Q. He, "Autoencoder and Its Various Variants," 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 2018, pp. 415-419, doi: <https://doi.org/10.1109/SMC.2018.00080> (date of access: 29.11.2024).
45. Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. Proceedings of the 25th International Conference on Machine Learning, 1096–1103. Retrieved from <https://www.cs.toronto.edu/~larocheh/publications/icml-2008-denoising-autoencoders.pdf> (date of access: 29.11.2024).
46. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of Machine Learning Research, 11, 3371–3408. Retrieved from <https://jmlr.org/papers/volume11/vincent10a/vincent10a.pdf> (date of access: 29.11.2024).
47. Gao, L., Dupré la Tour, T., Tillman, H., Goh, G., Troll, R., Radford, A., Sutskever, I., Leike, J., & Wu, J. (2024). Scaling and evaluating sparse autoencoders. arXiv. <https://arxiv.org/abs/2406.04093> (date of access: 29.11.2024).
48. Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. *arXiv*. <https://arxiv.org/abs/1312.6114> (date of access: 29.11.2024).
49. Jerome H. Friedman. "Greedy function approximation: A gradient boosting machine.." *Ann. Statist.* 29 (5) 1189 - 1232, October 2001. <https://doi.org/10.1214/aos/1013203451> (date of access: 29.11.2024).
50. Hemashreekilari. Understanding Gradient Boosting. *Medium*. URL: <https://medium.com/@hemashreekilari9/understanding-gradient-boosting-632939b98764> (date of access: 29.11.2024).

51. He, Zhiyuan, et al. "Gradient Boosting Machine: A Survey." arXiv preprint arXiv:1908.06951, 2019. URL: <https://arxiv.org/pdf/1908.06951> (date of access: 29.11.2024).
52. Sigrist, F. (2020). Gradient and Newton Boosting for Classification and Regression. arXiv preprint arXiv:1808.03064. URL: <https://doi.org/10.48550/arXiv.1808.03064> (date of access: 29.11.2024).
53. Catboost vs Xgboost vs Lightgbm: Which is Best? – Grid Dynamics. Grid Dynamics. URL: <https://www.griddynamics.com/blog/xgboost-vs-catboost-vs-lightgbm> (date of access: 29.11.2024).
54. Erdebilli, Babek, and Burcu Devrim-İçtenbaş. 2022. "Ensemble Voting Regression Based on Machine Learning for Predicting Medical Waste: A Case from Turkey" Mathematics 10, no. 14: 2466. <https://doi.org/10.3390/math10142466> (date of access: 29.11.2024).
55. Chen, S., & Nguyen, M. L. (2022). RRMSE Voting Regressor: A weighting function based improvement to ensemble regression. arXiv preprint arXiv:2207.04837. Available at <https://arxiv.org/abs/2207.04837> (date of access: 29.11.2024).
56. Afendras, G., & Markatou, M. (2015). On the optimality of the training/test size and resampling efficacy of cross-validation estimators of generalization error. arXiv:1511.02980 [math.ST]. URL: <https://doi.org/10.48550/arXiv.1511.02980> (date of access: 29.11.2024).
57. Moss, G. B., Leslie, D. S., & Rayson, P. (2018). Using J-K-fold cross validation to reduce variance when tuning NLP models. arXiv:1806.07139 [cs.CL]. URL: <https://doi.org/10.48550/arXiv.1806.07139> (date of access: 29.11.2024).
58. Brownlee, J. (2020). A Gentle Introduction to Threshold-Moving for Imbalanced Classification. Machine Learning Mastery. Available at: <https://machinelearningmastery.com/threshold-moving-for-imbalanced-classification/> (date of access: 29.11.2024).
59. Lugiez, F., Guimaraes, B., & Abreu, M. (2014). Threshold optimization for classification in imbalanced data in a high-energy physics application. Advances

in Data Analysis and Classification, 8(3), 253–276. DOI:

<https://doi.org/10.1007/s11634-014-0167-5> (date of access: 29.11.2024).

60. Healthy Brain Network - The Healthy Brain Network is a landmark mental health study that will help children in New York City and around the world. URL:

<https://healthybrainnetwork.org/> (date of access: 29.11.2024).



Звіт подібності

метадані

Заголовок

Василенко_Іщук_плагіат

Автор

Науковий керівник / Експерт

Василенко

Іщук

підрозділ

кафедра математичного моделювання та статистики

Тривога

У цьому розділі ви знайдете інформацію щодо текстових спотворень. Ці спотворення в тексті можуть говорити про МОЖЛИВІ маніпуляції в тексті. Спотворення в тексті можуть мати навмисний характер, але частіше характер технічних помилок при конвертації документа та його збереженні, тому ми рекомендуємо вам підходити до аналізу цього модуля відповідально. У разі виникнення запитань, просимо звертатися до нашої служби підтримки.

Заміна букв	Б	1
Інтервали	A→	0
Мікропробіли	.	14
Білі знаки	Б	0
Парафрази (SmartMarks)	a	9

Обсяг знайдених подібностей

Коефіцієнт подібності визначає, який відсоток тексту по відношенню до загального обсягу тексту було знайдено в різних джерелах. Зверніть увагу, що високі значення коефіцієнта не автоматично означають плагіат. Звіт має аналізувати компетентна / уповноважена особа.



КП 1

25

Довжина фрази для коефіцієнта подібності 2



КП 2

16518

Кількість слів



КЦ

131941

Кількість символів

Подібності за списком джерел

Нижче наведений список джерел. В цьому списку є джерела із різних баз даних. Колір тексту означає в якому джерелі він був знайдений. Ці джерела і значення Коефіцієнту Подібності не відображають прямого плагіату. Необхідно відкрити кожне джерело і проаналізувати зміст і правильність оформлення джерела.

10 найдовших фраз

Колір тексту

ПОРЯДКОВИЙ НОМЕР	НАЗВА ТА АДРЕСА ДЖЕРЕЛА URL (НАЗВА БАЗИ)	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)	
1	https://visnyk.chnpu.edu.ua/wp-content/uploads/2022/09/Khovrich-M.-Mekshun-A.pdf	20	0.12 %
2	https://ir.kneu.edu.ua/bitstreams/5b4feb2a-b118-4a75-9401-49db9f0703ee/download	16	0.10 %
3	https://ouyanglab.com/singlecell/references.html	14	0.08 %
4	https://dsum.edu.ua/uk/educational-programs/upep	13	0.08 %
5	http://dspace.luguniv.edu.ua/xmlui/bitstream/handle/123456789/3944/%E2%84%96%201%28324%29%2003%202019_%D0%A7.1.pdf?sequence=1&isAllowed=y	11	0.07 %

6	Висновки_Осетров.docx 12/13/2021 Kharkiv State Academy of Design and Fine Arts (KSADA)	10	0.06 %
7	http://lib.pnu.edu.ua:8080/bitstream/123456789/20571/1/%D0%9C%D0%B0%D0%B3%D1%96%D1%81%D1%82%D0%B5%D1%80%D1%81%D1%8C%D0%BA%D0%B0%20%D0%BE%D1%81%D0%BD%D0%BE%D0%B2%D0%BD%D0%B0%20%282%29%20%282%29.docx	10	0.06 %
8	КОМПЛЕКС ЗАХОДІВ ДЛЯ СТАБІЛІЗАЦІЇ ФІНАНСОВОЇ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА ПРИ ЗАГРОЗІ БАНКРУТСТВА І.С. Усатенко, О.П. Заборовець;	9	0.05 %
9	Моделювання кредитного ризику позичальників комерційного банку 2/11/2020 Kyiv National Economic University named after Vadym Hetman KNEU (кафедра економіко-математичного моделювання)	7	0.04 %
10	https://ir.kneu.edu.ua/bitstreams/077d7d50-85b4-4595-93d5-25ac1b503ae4/download	7	0.04 %

з бази даних RefBooks (0.05 %)

ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
джерело: Paperity		
1	КОМПЛЕКС ЗАХОДІВ ДЛЯ СТАБІЛІЗАЦІЇ ФІНАНСОВОЇ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА ПРИ ЗАГРОЗІ БАНКРУТСТВА І.С. Усатенко, О.П. Заборовець;	9 (1) 0.05 %

з домашньої бази даних (0.04 %)

ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
1	Моделювання кредитного ризику позичальників комерційного банку 2/11/2020 Kyiv National Economic University named after Vadym Hetman KNEU (кафедра економіко-математичного моделювання)	7 (1) 0.04 %

з програми обміну базами даних (0.06 %)

ПОРЯДКОВИЙ НОМЕР	ЗАГОЛОВОК	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
1	Висновки_Осетров.docx 12/13/2021 Kharkiv State Academy of Design and Fine Arts (KSADA)	10 (1) 0.06 %

з Інтернету (0.68 %)

ПОРЯДКОВИЙ НОМЕР	ДЖЕРЕЛО URL	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
1	https://visnyk.chnpu.edu.ua/wp-content/uploads/2022/09/Khovrich-M.-Mekshun-A.pdf	20 (1) 0.12 %
2	https://ir.kneu.edu.ua/bitstreams/5b4feb2a-b118-4a75-9401-49db9f0703ee/download	16 (1) 0.10 %
3	https://www.doccity.com/ru/kursovaya-po-psihologii-7/7279450/	16 (3) 0.10 %
4	https://ouyanglab.com/singlecell/references.html	14 (1) 0.08 %
5	https://dsum.edu.ua/uk/educational-programs/upep	13 (1) 0.08 %

6	http://dspace.luguniv.edu.ua/xmlui/bitstream/handle/123456789/3944/%E2%84%96%201%28324%29%2003%202019_%D0%A7.1.pdf?sequence=1&isAllowed=y	11 (1)	0.07 %
7	http://lib.pnu.edu.ua:8080/bitstream/123456789/20571/1/%D0%9C%D0%B0%D0%B3%D1%96%D1%81%D1%82%D0%B5%D1%80%D1%81%D1%8C%D0%BA%D0%B0%20%D0%BE%D1%81%D0%BD%D0%BE%D0%B2%D0%BD%D0%B0%20%282%29%20%282%29.docx	10 (1)	0.06 %
8	https://ir.kneu.edu.ua/bitstreams/077d7d50-85b4-4595-93d5-25ac1b503ae4/download	7 (1)	0.04 %
9	https://ir.kneu.edu.ua/bitstreams/bab2477b-8701-4c58-bbbd-7d6ee35d3fda/download	6 (1)	0.04 %

Список прийнятих фрагментів (немає прийнятих фрагментів)

ПОРЯДКОВИЙ НОМЕР	ЗМІСТ	КІЛЬКІСТЬ ОДНАКОВИХ СЛІВ (ФРАГМЕНТІВ)
------------------	-------	---------------------------------------

Василенко Д.П.

Електронна пошта: d.p.vasilenko@gmail.com

Науковий керівник : к.е.н, доцент Іщук Я.В.

Електронна пошта: ishchuk.yaroslava@kneu.edu.ua

Київський національний економічний університет імені Вадима Гетьмана
ОГЛЯД МЕТОДІВ DATA SCIENCE ДЛЯ ПРОГНОЗУВАННЯ СХИЛЬНОСТІ ДО
ІНТЕРНЕТ ЗАЛЕЖНОСТІ

Vasylenko D.P.

REVIEW OF DATA SCIENCE METHODS FOR PREDICTING INTERNET
ADDICTION

У роботі проаналізовані методичні підходи до проблеми інтернет-залежності серед молоді та сучасні методи її прогнозування. Акцентовано увагу на моделях машинного навчання, які дозволяють більш ефективно аналізувати фактори, що впливають на розвиток інтернет-залежності, та забезпечують точніші оцінки рівня залежності серед студентів.

Methodical approaches to the problem of Internet addiction among young people and modern methods of its prediction are analyzed in the work. Attention is focused on machine learning models, which allow more effective analysis of factors affecting the development of Internet addiction and provide more accurate assessments of the level of addiction among students.

Інтернет-залежність є актуальною проблемою сучасного суспільства, особливо серед молоді, оскільки надмірне використання інтернету негативно впливає на психічне здоров'я, соціальні навички та академічну успішність. Це також має значний економічний вплив, адже зниження продуктивності та зростання витрат на охорону здоров'я через психічні розлади призводять до додаткових витрат для суспільства та окремих інституцій.[1] У зв'язку з цим зростає потреба в нових підходах до діагностики та попередження інтернет-залежності. Використання методів машинного навчання для виявлення схильності до інтернет-залежності є перспективним напрямом, оскільки ці методи дозволяють більш точно аналізувати поведінкові дані та прогнозувати ризики залежності. Це робить їх важливим інструментом не лише для збереження психічного здоров'я молоді, але й для зменшення економічних втрат, пов'язаних із залежністю.

У статті "A new model of Internet addiction tendency based on machine learning" [2], представлений на Міжнародній конференції з комп'ютерних наук і технологій зв'язку (ICCSCT 2022) у Пекіні, Чжицзун Го та Лайган Чжан досліджують проблему інтернет-залежності та пропонують нову модель для аналізу багатфакторного впливу на схильність до неї.

У своєму дослідженні автори побудували багатфакторну вагову модель аналізу схильності до інтернет-залежності (IDT), використовуючи метод машинного навчання. Для досягнення цієї мети було обрано глибинну нейронну мережу Deep Belief Network (DBN), яка здатна визначати більш точний розподіл ваг

для різних факторів, що впливають на схильність до залежності. Через тренування цієї мережі модель надає можливість ідентифікувати ефективніші характеристики різних факторів, що призводить до високої точності прогнозування.

Автори також провели експериментальне порівняння між запропонованою моделлю та існуючими методами аналізу. Результати показали, що нова модель має переваги як у точності, так і в ефективності визначення факторів ризику, що можуть призводити до інтернет-залежності. Наприклад, порівняння з попередніми методами показало, що використання DBN для визначення ваг у багатфакторній моделі значно покращує результати прогнозування.[3]

Ця робота є цінним вкладом у галузі профілактики та дослідження інтернет-залежності. Запропонована модель дозволяє створити більш точну ідентифікацію факторів, що підвищують ризик розвитку залежності, і може слугувати основою для майбутніх досліджень та розробки інтервенційних заходів, які б сприяли здоровому використанню Інтернету. Дослідження Го та Чжана надає практичні рекомендації для фахівців, які займаються профілактикою залежностей, а також для психологів, які можуть використовувати отримані результати для розробки психологічних та поведінкових стратегій щодо попередження інтернет-залежності.

Інший підхід до вивчення цього питання запропонували китайські вчені Шуанг Чжан та Хуісі Ю. У статті "Online addiction analysis and identification of students by applying gd-LSTM algorithm to educational behaviour data," [4] опублікованій у журналі Journal of Intelligent Systems у березні 2024 року, автори досліджують проблему інтернет-залежності серед студентів. Вони пропонують новий підхід для оцінки та виявлення рівня залежності студентів, базуючись на аналізі поведінкових даних за допомогою алгоритму gd-LSTM, що комбінує методи Dropout і LSTM (Long Short-Term Memory).

Для виявлення студентів з інтернет-залежністю автори використовували стохастичні моделі (SM) для збору та аналізу поведінкових даних студентів. Було розроблено мережу з використанням gd-LSTM, яка дозволяє обробляти час-серійні особливості поведінки студентів. Ця модель фіксує багатовимірні вектори даних для точнішого прогнозування та оцінки рівня залежності.[5] В експерименті, що охоплював 5,861 студента, було виявлено, що 4.23% з них мають ознаки інтернет-залежності, при цьому 95.66% залежних були чоловіками.

Загалом, модель оцінювала інтернет-залежність за чотирма вимірами, а результати прогнозів у цих вимірах (з оцінками 0.73, 0.72, 0.74 та 0.74) продемонстрували, що gd-LSTM має переваги у порівнянні з іншими моделями, такими як LSTM та CNN. Це підтверджує, що запропонований алгоритм є більш ефективним у забезпеченні точності та рівномірності оцінок.

Автори дійшли висновку, що їхня модель може слугувати інструментом оцінки ризиків інтернет-залежності серед студентів, однак вона не є діагностичним засобом у медичному розумінні. gd-LSTM модель є важливим кроком у дослідженні поведінкових особливостей студентів і може бути застосована для оцінки інтернет-залежності в масштабних освітніх установах. Підхід дозволяє отримати більш точну

картину залежності та може бути корисним для розробки інтервенційних програм, спрямованих на підтримку психічного здоров'я та навчальної успішності студентів.

Література:

1. Cash, H., Rae, C. D., Steel, A. H., & Winkler, A. (2012). Internet Addiction: A Brief Summary of Research and Practice. *Current psychiatry reviews*, 8(4), 292–298. <https://doi.org/10.2174/157340012803520513>
2. Zhijun Guo, Laigang Zhang. A new model of Internet addiction tendency based on machine learning// Third International Conference on Computer Science and Communication Technology (ICCSCT 2022). 2022. Proceedings Volume 12506. №125061X. URL: <https://doi.org/10.1117/12.2662545> (date: 28.12.2022).
3. Hua, Yuming & Guo, Junhai & Zhao, Hua. (2015). Deep Belief Networks and deep learning. *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things, ICIT 2015*. 1-4. 10.1109/ICAIOT.2015.7111524. URL: https://www.researchgate.net/publication/282318042_Deep_Belief_Networks_and_deep_learning
4. Shuang Zhang, Huisi Yu. Online addiction analysis and identification of students by applying gd-LSTM algorithm to educational behaviour data // *Journal of Intelligent Systems*. 2024. T. 33. №1. C. 20230102. URL: <https://doi.org/10.1515/jisys-2023-0102>.
5. W. Zhang, H. Quan, W. Zhang and Z. Li, "Short-Term Wind Power Interval Prediction Based on GD-LSTM and Bootstrap Techniques," *2022 IEEE 5th International Electrical and Energy Conference (CIEEC)*, Nangjing, China, 2022, pp. 2626-2631. URL: <https://ieeexplore.ieee.org/document/9846316>.