КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
Економічний факультет
Кафедра економічної кібернетики

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА
« ПРОГНОЗУВАННЯ ВІДТОКУ КЛІЄНТІВ ЗА ДОПОМОГОЮ ГІБРИДНОЇ МОДЕЛІ НЕЙРОННИХ МЕРЕЖ »

студента 4 курсу
спеціальності 051 «Економіка»
ОПП «Економічна кібернетика»
денної форми навчання
Мордатенка Олександра
Костянтиновича

Науковий керівник:
к. ф.-м. н., доцент Кравець
Тетяна Вікторівна

Засвідчую, що у цій дипломній роботі немає запозичень із праць інших авторів без відповідних посилань

Студент_____(підпис)

Роботу допущено до захисту перед ЕК
рішенням кафедри економічної кібернетики
від 12.06 2025р., протокол № 15
Завідувач кафедри:
доктор економічних наук, професор
Ляшенко Олена Ігорівна

_____
(підпис)

КИЇВ – 2025

РЕФЕРАТ

Кваліфікаційна робота бакалавра містить: 57 ст., 16 рис., 3 табл., 26 джерел, 1 додаток.

Ключові слова: класифікація, відтік клієнтів, моделювання, прогнозування, машинне навчання, глибоке навчання, гібридна модель.

Об'єкт дослідження: процес взаємодії клієнтів із продуктами або послугами комерційних компаній, що призводить до явища відтоку клієнтів.

Предмет дослідження: моделі класифікації для прогнозування відтоку клієнтів на основі нейронних мереж.

Мета дослідження: побудувати гібридну модель класифікації для прогнозування відтоку клієнтів й порівняти її з результатами інших загальновідомих моделей.

Методи дослідження: системний підхід, аналіз та синтез, індукція і дедукція, абстрагування, порівняння та узагальнення, статистичний аналіз, економіко-математичне моделювання.

Наукова новизна, теоретична значимість дослідження: проведено детальний аналіз явища відтоку клієнтів, чинників впливу на дане явища та можливість впливати на них за допомогою прогнозування цього явища; побудовано гібридну модель глибокого навчання для прогнозування відтоку клієнтів й проведене порівняння якості даної моделі з вже загальновідомими підходами.

Практична цінність: можливість перевикористання результатів кваліфікаційної роботи для створення гібридної моделі прогнозування відтоку клієнтів переважно у галузях, що базуються на підписці або регулярних покупках.

RESUME

Taras Shevchenko National University of Kyiv,
Faculty of Economics, Department of Economic Cybernetics.

Key words: classification, customer churn, modeling, forecasting, machine learning, deep learning, hybrid model.

The graduation thesis: Customer churn prediction using a hybrid neural network model.

The practical value: the possibility of reusing the results of qualification work to create a hybrid model for predicting customer churn, mainly in subscription-based or regular purchase industries.

Pages – 57, figures – 16, tables – 3, references – 26, append. – 1.

ЗМІСТ

ВСТУП 5

РОЗДІЛ 1. ВИЗНАЧЕННЯ ВІДТОКУ КЛІЄНТІВ, ЙОГО ВПЛИВ НА БІЗНЕС 9

1.1. Визначення й особливості явища відтоку клієнтів 9

1.2. Основні метрики для вимірювання відтоку 12

1.3. Прямі й непрямі економічні наслідки відтоку клієнтів для бізнесу 14

РОЗДІЛ 2. ПРОГНОЗУВАННЯ ВІДТОКУ КЛІЄНТІВ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ: АКТУАЛЬНІСТЬ ТА ЗАСТОСУВАННЯ 18

2.1. Нейронні мережі як альтернатива традиційним моделям прогнозування 18

2.2. Огляд базових моделей нейронних мереж 20

2.3. Опис гібридної моделі A-BiLSTM-CNN-Net 25

РОЗДІЛ 3. ПРАКТИЧНЕ ЗАСТОСУВАННЯ ГІБРИДНОЇ МОДЕЛІ ДЛЯ ПРОГНОЗУВАННЯ ВІДТОКУ КЛІЄНТІВ 30

3.1. Огляд датасету 30

3.2. Підготовка даних та тренування моделі 38

3.3. Оцінка якості гібридної моделі 39

ВИСНОВКИ 44

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 46

ДОДАТКИ 50

ВСТУП

В умовах сучасної цифрової економіки, що характеризується високою конкуренцією та насиченістю ринків, боротьба за клієнта стає ключовим фактором успіху будь-якого бізнесу. Тож економічна доцільність утримання існуючих клієнтів є беззаперечною.

Залучення нового споживача коштує в середньому в 5-7 разів дорожче, ніж збереження лояльного [1]. Відтік клієнтів напряму впливає на дохід, прибутковість та довгострокову стабільність компанії. Тому розробка ефективних інструментів для завчасного прогнозування та запобігання відтоку є однією з найпріоритетніших задач для бізнесу в таких галузях, як телекомунікації [5], банківська справа, електронна комерція, SaaS (Software as a service) [4] та стрімінгові сервіси.

Проблема прогнозування відтоку не є новою. Для її розв'язання успішно застосовуються класичні статистичні підходи (наприклад, логістична регресія) та більш сучасні методи машинного навчання (градієнтний бустинг, random forest). Ці методи добре працюють зі структурованими, табличними даними і часто забезпечують високу точність та інтерпретованість результатів. Однак вони мають обмеження при роботі з великими масивами складних, нелінійних та, що особливо важливо, послідовних даних, які характеризують поведінку сучасного клієнта (історія переглядів, послідовність кліків, патерни використання мобільних додатків, текстові звернення до служби підтримки). [19]

Саме тут на перший план виходять нейронні мережі (НМ), особливо їхні глибокі архітектури, які пропонують фундаментальні переваги: [3]

- нейронні мережі здатні автоматично виявляти приховані, неочевидні патерни в даних, які неможливо ідентифікувати за допомогою лінійних моделей;
- рекурентні нейронні мережі та їхні вдосконалені версії, як-от моделі довгої короткочасної пам’яті, спеціально створені для аналізу часових послідовностей, дозволяючи моделювати еволюцію поведінки клієнта в часі [2].

Сьогодні наукова проблема полягає вже не в тому, чи використовувати нейронні мережі для прогнозування відтоку, а в тому, які саме архітектури та підходи є найефективнішими для вирішення конкретних бізнес-задач. Простих багатошарових перцептронів [21] вже недостатньо для конкурентної переваги. Сучасні дослідження зосереджені на більш складних підходах, таких як: гібридні моделі, удосконалені архітектури для послідовностей, інтерпретованість та пояснюваність

Для України, з її висококонкурентними ринками телекомунікацій, банківських послуг, IT та ритейлу, а також в умовах повоєнного відновлення економіки, ефективне утримання клієнтської бази набуває стратегічного значення. Втрата кожного клієнта — це не лише втрата доходу, а й потенційне ослаблення позицій вітчизняного бізнесу.

Таким чином, актуальність дослідження полягає у переході від загальних моделей прогнозування до створення спеціалізованих та, що найважливіше, дієвих інструментів на основі передових нейромережевих архітектур. Це є не лише важливим науковим завданням, а й нагальною вимогою для сталого розвитку української економіки й виходу українських компаній на нові західні ринки.

Об'єкт дослідження: процес взаємодії клієнтів із продуктами або послугами комерційних компаній, що призводить до явища відтоку клієнтів, переважно у галузях, що базуються на підписці або регулярних покупках.

Предмет дослідження: методи, моделі та алгоритми прогнозування відтоку клієнтів на основі нейронних мереж.

Мета дослідження: перевірити перспективи та ефективність нового гібридного підходу для прогнозування відтоку клієнтів шляхом розробки та валідації комплексної моделі на основі глибоких нейронних мереж, здатної аналізувати складні нелінійні та послідовні поведінкові патерни.

Завдання дослідження:

- проаналізувати проблему відтоку клієнтів, її економічний вплив на бізнес та розглянути актуальність застосування саме нейронних мереж

для прогнозування відтоку, порівнявши її результати з традиційними методами машинного навчання;

- дослідити теоретичні основи та архітектури різних типів нейронних мереж, визначивши їхні сильні та слабкі сторони в контексті аналізу клієнтських даних;
- проаналізувати датасет, згідно з яким, виокремити основні причини відтоку клієнтів;
- розробити архітектуру гібридної моделі, що ефективно поєднує аналіз статичних та послідовних даних для комплексного розуміння поведінки клієнта;
- реалізувати розроблену модель з використанням сучасних програмних фреймворків;
- оцінити ефективність моделі за рядом ключових метрик, порівнюючи її результати з базовими моделями машинного навчання;
- сформулювати практичні рекомендації для бізнесу щодо впровадження розробленої моделі та розробки цільових стратегій утримання клієнтів на основі отриманих результатів.

Робота складається з 3-х розділів та висновку:

- В І розділі описується об’єкт дослідження, його характеристики, особливості, чинники впливу. Наведено основні тенденції ринку та метрики вимірювання руху користувачів.
- В ІІ розділі представлений предмет дослідження - методи та моделі, за допомогою яких здійснюється прогнозування; наведена архітектура представленої гібридної моделі.
- В ІІІ розділі представлена практична частина: огляд датасету, його аналіз, обробка даних, прогнозування та порівняння результатів з іншими моделями.
- У висновку наведено основні результати проведеного дослідження.

Увесь процес програмно описаний за допомогою Python [16] коду та використанням ряду фреймворків. Для передобробки даних, валідації та

розрахунку метрик точности застосовується Scikit-learn [16]. А побудова, компіляція та тренування нейронної мережі здійснюється за допомогою TensorFlow [15] та його високорівневим API Keras [17], що є центральним фреймворком, відповідальним за створення, налаштування та навчання гібридної моделі глибокого навчання.

РОЗДІЛ 1. ВИЗНАЧЕННЯ ВІДТОКУ КЛІЄНТІВ, ЙОГО ВПЛИВ НА БІЗНЕС

1.1. Визначення й особливості явища відтоку клієнтів

Відтік клієнтів, також відомий як коефіцієнт відтоку клієнтів (Customer Churn Rate, CR), є показником, що відображає кількість або відсоток клієнтів, які припинили користуватися послугами чи продуктами компанії протягом певного, чітко визначеного періоду. [6] Цей показник слугує важливим індикатором ефективності бізнес-діяльності, особливо в контексті здатності компанії утримувати своїх клієнтів та підтримувати їхню лояльність.

Важливо розуміти, що відтік – це не просто статистична втрата клієнта. Високий рівень відтоку часто сигналізує про наявність глибших, системних проблем у бізнесі. Це може бути ознакою невідповідності продукту чи послуги очікуванням ринку, проблем із ціновою політикою, низької якості обслуговування клієнтів або ж свідченням того, що конкуренти пропонують більш привабливі умови. Якщо спостерігається масовий відхід клієнтів, це рідко буває випадковістю; скоріше, це кумулятивний ефект, що вказує на системні недоліки в бізнес-моделі, самому продукті або в організації клієнтського сервісу. Наприклад, незадоволеність клієнтів або невірно обране маркетингове позиціонування можуть бути одними з ключових драйверів високого відтоку.

Для ефективного аналізу та управління відтоком важливо розрізняти його типи, оскільки кожен з них має свої специфічні причини та вимагає різних підходів до вирішення.

Добровільний відтік – це тип, який відбувається, коли клієнт свідомо та з власної ініціативи приймає рішення припинити співпрацю з компанією; таке рішення часто є наслідком глибших проблем, пов'язаних із ціннісною пропозицією продукту чи послуги, або ж свідчить про незадоволеність рівнем обслуговування. Вимушений – відтік, який відбувається поза контролем або бажанням самого клієнта; найчастіше він пов'язаний із проблемами з оплатою (наприклад, закінчення терміну дії банківської картки, недостатність коштів на

рахунку) або певними сервісними проблемами з боку компанії. У випадку пасивного відтоку клієнт поступово припиняє взаємодію з компанією та її продуктами/послугами без формального розірвання відносин, щ може свідчити про брак залученості клієнта, втрату інтересу або про те, що корисність продукту для нього стала неочевидною. Відтік доходу – це тип відтоку, котрий відображає втрату регулярного доходу, що виникає через скасування клієнтами підписок, відмову від послуг або їх перехід на дешевші тарифні плани; такий відтік надає фінансову перспективу на втрати, показуючи не просто кількість втрачених клієнтів, а й обсяг втрачених грошей.

Для адекватної оцінки власних показників відтоку, компаніям корисно порівнювати їх із середньогалузевими значеннями. Однак варто пам'ятати, що ці цифри є орієнтовними, і "нормальний" рівень відтоку може суттєво відрізнятися залежно від специфіки бізнесу, його моделі та стадії розвитку.

Згідно з даними на квітень 2024 року, середні показники річного відтоку для деяких галузей такі: [7]

- Енергетика та комунальні послуги: 11%
- Споживчі товари та роздрібна торгівля: 5.5%
- Виробництво: 35%
- Комп'ютерне програмне забезпечення: 14%
- SaaS (Software as a Service): 4.67%
- IT-послуги: 12%
- Професійні послуги: 27%
- Телекомунікації: 31%

Для SaaS-сектору показники можуть варіюватися: 3-5% щомісячного відтоку для компаній, орієнтованих на малий та середній бізнес (SMB), 1-2% для компаній середнього ринку, і менше 2% для великих корпоративних клієнтів (Enterprise). B2C SaaS компанії часто мають вищий щомісячний відтік, в діапазоні 3-7%. Для eCommerce бізнесів, що працюють за моделлю підписки, щомісячний відтік може становити 5-10%, а для медіа та стрімінгових сервісів – 6-8%. Фінансовий сектор демонструє середньорічний відтік близько 19%.

Загалом, деякі джерела вважають "хорошим" середнім показником відтоку по всіх галузях близько 4%, хоча ідеальний показник – 0% – є практично недосяжним [8] [9]. Інші дослідження вказують, що середній річний показник відтоку на рівні 5-7% може вважатися нормою для певних ринків.

Порівняння власного показника відтоку із середньогалузевими є важливим, але не єдиним критерієм оцінки. Необхідно враховувати стадію розвитку компанії (стартапи можуть мати вищий відтік на початкових етапах), бізнес-модель (наприклад, B2B зазвичай має нижчий відтік, ніж B2C, а моделі підписки більш чутливі до відтоку, ніж разові продажі) та специфіку цільової аудиторії. "Нормальний" відтік є відносним поняттям, і компанія повинна прагнути до його поступового зниження, орієнтуючись не тільки на загальновизначені критерії, але й на власну динаміку та економічну доцільність.

Таблиця 1.1

Середні показники відтоку клієнтів за галузями

Галузь	**Середній показник відтоку (%),** станом на квітень 2024 року [7] [10]
SaaS	4.67 (річний)
SaaS (SMB-focused)	3-5 (місячний)
SaaS (Enterprise B2B)	<2 (місячний)
SaaS (B2C)	3-7 (місячний)
Роздрібна торгівля та споживчі товари	5.5 (річний)
eCommerce (підписка)	5-10 (місячний)
Телекомунікації	31 (річний)
Енергетика та комунальні послуги	11 (річний)

Продовження табл. 1.1

Галузь	**Середній показник відтоку (%), станом на квітень 2024 року**
Виробництво	35 (річний)
Комп'ютерне програмне забезпечення	14 (річний)
IT-послуги	12 (річний)
Професійні послуги	27 (річний)
Медіа та стрімінг	6-8 (місячний)
Фінансові послуги	~19 (річний)

Джерело: [7] [10]

1.2. Основні метрики для вимірювання відтоку

Для кількісної оцінки відтоку клієнтів та його впливу на бізнес використовується низка ключових метрик таких як CR, MRR Churn, Retention Rate, ACL, LTV. Розгляньмо детальніше кожну з них.

Коефіцієнт відтоку клієнтів (Customer Churn Rate, CR) – це базовий показник, що відображає відсоток клієнтів, яких компанія втратила за певний період. Існує кілька підходів до його розрахунку:

1. *Базова формула*:

$$CR = \frac{\text{кількість втрачених клієнтів}}{\text{загальна кількість клієнтів на початок періоду}} * 100\%$$

2. *Формула з урахуванням зміни загальної кількості клієнтів*:

$$CR = \frac{C_1 - C_2}{C_1} * 100\%, \qquad (1.1)$$

де C_1 – кількість клієнтів на початок періоду;

C_2 – кількість клієнтів на кінець періоду.

Цей показник може бути негативним, якщо приріст клієнтської бази перевищує втрати, що свідчить про загальне зростання.

3. *Формула для оцінки утримання існуючих клієнтів* ідентична до формули (1.1), але в цьому випадку C_2 означає кількість клієнтів, що залишилися з тих, хто був на початку періоду (нові клієнти не враховуються). Коефіцієнт відтоку за цією формулою не може бути негативним

4. *Формула з урахуванням залучених нових клієнтів:*

$$CR = \frac{C_1 + C_3 - C_2}{C_1} * 100\%, \qquad (1.2)$$

де C_1 – кількість клієнтів на початок періоду;

C_2 – кількість клієнтів на кінець періоду;

C_3 – кількість нових клієнтів на кінець періоду.

Ця формула дає більш реалістичну картину щодо ефективності утримання існуючої бази, враховуючи приплив нових клієнтів

Коефіцієнт відтоку щомісячного регулярного доходу (Monthly Recurring Revenue Churn, MRR Churn) – це метрика, яка показує відсоток щомісячного регулярного доходу (MRR), втраченого через скасування підписок або перехід клієнтів на дешевші тарифні плани. Цей показник не враховує дохід, отриманий від розширення співпраці з існуючими клієнтами (апсейли, крос-сейли). Існує валовий та чистий MRR Churn. Валовий – не враховує дохід, отриманий від розширення співпраці з існуючими клієнтами (апсейли, крос-сейли), а чистий – навпаки, що свідчить про те, що він може бути негативним.

Коефіцієнт утримання клієнтів (Customer Retention Rate, CRR) – метрика, що відображає відсоток клієнтів, яких компанії вдалося утримати протягом певного періоду. Розраховується як $CRR = 100\% - CR$ (якщо CR виражений у відсотках) або $CRR = 1 - CR$ (якщо CR виражений у частках).

Середній термін життєвого циклу клієнта (Average Customer Lifetime, ACL) – метрика, яка показує середній час, протягом якого клієнт залишається активним користувачем послуг або продуктів компанії. Розраховується за

формулою $ACL = 1/CR$ (якщо CR виражений у частках за відповідний період, наприклад, місяць) або $ACL = 100/CR$ (якщо CR виражений у відсотках).

Цінність клієнта протягом життєвого циклу в продукті (Customer Lifetime Value, LTV або CLV) – це прогноз загального прибутку, який компанія очікує отримати від одного клієнта за весь час його взаємодії з бізнесом. Важливо, щоб LTV був значно вищим за вартість залучення клієнта (CAC).

Вибір конкретної формули для розрахунку коефіцієнта відтоку клієнтів (CR) залежить від мети аналізу. Якщо компанія прагне оцінити загальну динаміку своєї клієнтської бази, доцільно використовувати формулу, яка враховує нових клієнтів і може дати негативний результат (що свідчить про зростання бази). Однак, якщо основна мета – оцінити ефективність стратегій утримання саме існуючих клієнтів, слід застосовувати формули, які фокусуються на кількості втрачених клієнтів з початкової бази, і де CR не може бути негативним.

Особливо значущим індикатором здоров'я бізнесу є негативний чистий MRR Churn. Це означає, що зростання доходу від існуючих клієнтів (завдяки успішним апсейлам, крос-сейлам або переходу на дорожчі тарифи) перевищує фінансові втрати від клієнтів, що пішли. Такий показник свідчить про високу цінність продукту для лояльної клієнтської бази та ефективну стратегію монетизації існуючих клієнтів. Це не просто компенсація втрат, а активне зростання доходу за рахунок поглиблення та розширення відносин з наявною клієнтською базою.

1.3. Прямі й непрямі економічні наслідки відтоку клієнтів для бізнесу

Відтік клієнтів має глибокий та багатогранний економічний вплив на бізнес, що виходить далеко за межі простої втрати одного чи кількох замовників. Цей вплив можна умовно розділити на прямі та непрямі економічні наслідки. Прямі наслідки відтоку безпосередньо відображаються у фінансових звітах компанії та включають різні аспекти. Прямі наслідки відтоку безпосередньо відображаються у фінансових звітах компанії та включають втрату доходу, витрати на залучення нових клієнтів

Втрата доходу – це найбільш очевидний та негайний наслідок, оскільки кожен клієнт, що йде, забирає з собою потенційний дохід, який він міг би генерувати. Це особливо критично для бізнесів, що працюють за моделлю підписки або передбачають регулярні повторні покупки, де відтік безпосередньо зменшує щомісячний регулярний дохід (MRR) або річний регулярний дохід (ARR). Наприклад, гіпотетична компанія, що обслуговує 500 000 клієнтів із середнім річним доходом $250 на клієнта, при річному рівні відтоку 40% втрачає $50 мільйонів доходу щороку лише від прямих втрат. Розрахунок втраченого MRR, як валового, так і чистого, є ключовим для розуміння масштабів цієї проблеми.

Другий за важливістю фактор – це витрати на залучення нових клієнтів, порівняно з утриманням. Загальновизнаним фактом є те, що залучення нового клієнта обходиться значно дорожче, ніж утримання існуючого. Різні джерела вказують, що ця різниця може становити від 5 до 25 разів. Вартість залучення клієнта (CAC) розраховується як відношення загальних витрат на продажі та маркетинг за певний період до кількості нових клієнтів, залучених за той самий період. Тому високий рівень відтоку змушує компанію постійно спрямовувати значні ресурси на залучення нових клієнтів лише для того, щоб компенсувати втрати та підтримувати поточний рівень доходу, не кажучи вже про зростання.

Ну й також не варто забувати, що відтік клієнтів також генерує додаткові операційні витрати. Сюди можна віднести витрати, пов'язані з обробкою запитів на скасування послуг, процедурами повернення коштів або товарів, закриттям облікових записів клієнтів. Крім того, співробітники служби підтримки можуть витрачати значний час на спроби утримати клієнтів, що йдуть, або на обробку їхніх скарг, що відволікає ресурси від обслуговування лояльних клієнтів.

Прямі фінансові втрати від відтоку мають тенденцію до накопичення та посилення. Це не просто одноразова втрата доходу від одного конкретного клієнта, а постійний "витік" з дохідної бази компанії. Цей процес схожий на спробу наповнити діряве відро: компанія змушена докладати все більше зусиль (і витрачати більше коштів на маркетинг та продажі) лише для того, щоб

компенсувати постійні втрати, що значно ускладнює досягнення стабільного зростання та прибутковості.

Окрім прямих фінансових втрат, відтік клієнтів спричиняє низку непрямих економічних наслідків, які, хоч і важче піддаються кількісній оцінці, можуть мати навіть більш руйнівний та довгостроковий вплив на бізнес, зокрема: зниження LTV, втрата можливостей для додаткових та перехресних продажів, зниження довіри інвесторів та ринкової частки. Непрямі наслідки відтоку, хоча їх і важче виміряти в конкретних грошових одиницях у короткостроковій перспективі, можуть мати значно більш довгостроковий та руйнівний вплив на бізнес, ніж прямі втрати доходу. Пошкоджена репутація, втрата довіри клієнтів та інвесторів, демотивовані співробітники – це проблеми, вирішення яких потребує значно більше часу, зусиль та ресурсів, ніж просто залучення нових клієнтів для компенсації поточних втрат.

Більше того, ці фактори можуть створювати своєрідну петлю негативного зворотного зв'язку. Високий рівень відтоку призводить до погіршення репутації бренду, що, у свою чергу, ускладнює залучення нових якісних клієнтів. Це змушує компанію збільшувати витрати на маркетинг та продажі (CAC) для залучення хоч якихось клієнтів, що негативно впливає на прибутковість. Зниження прибутковості може призвести до скорочення інвестицій у розвиток продукту, покращення сервісу або навчання персоналу, що, своєю чергою, може спричинити подальше погіршення якості та, як наслідок, ще більший відтік клієнтів. Таким чином, відтік може запустити самопідсилювальний негативний цикл, з якого бізнесу буде дуже важко вибратися.

Точне прогнозування відтоку клієнтів надає компаніям хороший інструмент для того, аби мінімізувати втрати та зміцнити фінансову впевненість компанії. Головна перевага полягає в можливості ідентифікувати клієнтів, схильних до відтоку, ще до того, як вони фактично приймуть рішення піти. Це дозволяє перейти від реактивного підходу (спроб повернути клієнта, який вже пішов) до проактивного утримання.

Маючи інформацію про потенційних "відмовників", компанії можуть розробляти та впроваджувати цільові стратегії утримання. Це можуть бути персоналізовані пропозиції, спеціальні знижки, покращення умов обслуговування, надання додаткових освітніх матеріалів щодо використання продукту, або проактивне звернення служби підтримки для вирішення можливих проблем. Такі заходи не лише знижують ймовірність відтоку конкретного клієнта, але й підвищують його загальну лояльність та задоволеність.

Ефективне утримання клієнтів безпосередньо впливає на їхню цінність протягом життєвого циклу в продукті (LTV). Тобто чим довше клієнт залишається з компанією, тим більший дохід він генерує протягом усього періоду співпраці. Збільшення LTV, у свою чергу, сприяє стабілізації загального доходу компанії та покращує її фінансові показники. Навіть невелике зниження коефіцієнта відтоку може призвести до суттєвого зростання прибутку. Наприклад, за деякими оцінками, зниження відтоку на 5% може збільшити прибуток компанії на 25-85%. [11] Таким чином, інвестиції в системи прогнозування відтоку та подальші заходи з утримання є стратегічно виправданими та економічно необхідними.

РОЗДІЛ 2. ПРОГНОЗУВАННЯ ВІДТОКУ КЛІЄНТІВ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ: АКТУАЛЬНІСТЬ ТА ЗАСТОСУВАННЯ

2.1. Нейронні мережі як альтернатива традиційним моделям прогнозування

Традиційні методи аналізу та прогнозування відтоку клієнтів, такі як логістична регресія, дерева рішень або кластерний аналіз, відіграли важливу роль і продовжують використовуватися. Однак, вони мають певні обмеження, особливо в умовах сучасного бізнесу, що характеризується великими обсягами даних (Big Data), складними взаємозв'язками та динамічною поведінкою клієнтів.

Одним з основних недоліків традиційних підходів є їхня обмежена здатність виявляти складні нелінійні залежності в даних. Поведінка клієнта, що призводить до відтоку, рідко визначається одним або двома факторами; частіше це результат комплексної взаємодії багатьох змінних. Традиційні моделі можуть не вловлювати ці тонкі патерни. Крім того, багато з цих методів покладаються на статичні дані, тобто аналізують "знімок" стану клієнта на певний момент часу, і можуть недостатньо ефективно враховувати динаміку його взаємодії з продуктом або сервісом протягом часу. Процес інжинірингу ознак (feature engineering) для таких моделей часто вимагає значних ручних зусиль та глибокої експертизи в доменній області. До того ж класичні методи зазвичай дають змогу прогнозувати лише на рівні когорт, у той час як нейронні мережі дозволяють це робити на рівні користувача.

Саме тут на перший план виходять нейронні мережі, зокрема моделі глибокого навчання (Deep Learning). Нейронні мережі мають унікальну здатність автоматично вивчати складні патерни та ієрархічні представлення ознак безпосередньо з сирих даних. Це значно зменшує потребу в ручному інжинірингу ознак і дозволяє виявляти приховані, неочевидні залежності, які можуть бути пропущені традиційними методами. Завдяки своїй гнучкості та потужності,

нейронні мережі здатні обробляти великі та різнорідні набори даних, включаючи табличні дані, текстову інформацію (наприклад, відгуки клієнтів, звернення до служби підтримки) та послідовні дані (історія активності користувача). Це робить їх особливо перспективним інструментом для побудови більш точних та надійних моделей прогнозування відтоку клієнтів.

Актуальність використання нейронних мереж для прогнозування відтоку полягає не лише в потенційному підвищенні точності прогнозів. Це має під собою також і зміну в підході до управління взаємодії з клієнтами. Традиційно, аналіз відтоку часто був ретроспективним: компанії аналізували причини, чому клієнти вже пішли, і намагалися зробити висновки на майбутнє. Нейронні мережі, завдяки своїй прогнозній силі, дозволяють ідентифікувати клієнтів з групи ризику заздалегідь. Це уможливлює перехід від реактивного реагування на вже здійснений відтік до проактивного його запобігання. Такий зсув підходів дозволяє компаніям діяти на випередження, зберігаючи цінні клієнтські відносини та пов'язаний з ними дохід. Однак, цей перехід вимагає не тільки впровадження нових технологій, але й відповідної перебудови внутрішніх бізнес-процесів. Необхідна інтеграція прогнозних моделей у внутрішні системи, розробка та впровадження персоналізованих стратегій утримання, заснованих на отриманих прогнозах. Крім того, значно зростає важливість якості, повноти та своєчасності даних, що збираються про клієнтів, що також є викликом для інженерів тієї чи іншої компанії, аби доставляти дані в реальному часі. І часто проблема постає не через низьку обізнаність інженерів, а саме обмеження тих інструментів і джерел даних, з якими вони працюють.

Більше того, систематичний аналіз прогнозів відтоку та факторів, що на нього впливають (які стають більш прозорими завдяки сучасним методам інтерпретації нейронних мереж), слугує не лише для утримання конкретних клієнтів. Він надає цінний зворотний зв'язок для вдосконалення самого продукту чи послуги. Якщо модель прогнозує високий рівень відтоку для певного сегмента клієнтів або через певні характеристики продукту, це може сигналізувати про наявність системних проблем, таких як невідповідність продукту потребам

ринку, наявність критичних багів, недоліки в ціноутворенні або сервісі. Інтеграція таких інсайтів, отриманих з моделей прогнозування відтоку, в цикли розробки продукту (product development lifecycle) та в загальну маркетингову стратегію може призвести до створення більш конкурентоспроможних пропозицій, кращої відповідності ринковим потребам. Це, в свою чергу, сприятиме органічному зниженню відтоку в довгостроковій перспективі.

Таким чином, прогнозування відтоку за допомогою нейронних мереж перетворюється з локального інструменту утримання на важливий аспект для якісного розвитку та зростання бізнесу.

2.2. Огляд базових моделей нейронних мереж

Проста модель рекурентної нейронної мережі (RNN: recurrent neural network)

Рекурентна нейронна мережа (RNN) може бути визначена як особлива форма штучної нейронної мережі, яка була структурована для роботи з послідовними даними. [21] [24] Ці алгоритми глибокого навчання зазвичай використовуються для вирішення порядкових або часових завдань, таких як переклад мови, обробка природної мови (NLP), розпізнавання мови і підписи до зображень; вони включені в такі популярні програми, як Siri, голосовий пошук і Google Translate [19]. Вони відрізняються своєю "пам'яттю", оскільки беруть інформацію з попередніх входів, щоб впливати на поточний вхід і вихід. У той час як традиційні глибокі нейронні мережі припускають, що входи і виходи є незалежними один від одного, вихід рекурентних нейронних мереж залежить від попередніх елементів у послідовності. Хоча майбутні події також можуть бути корисними для визначення виходу певної послідовності, односпрямовані рекурентні нейронні мережі не можуть врахувати ці події у своїх прогнозах [14]. Рівняння мають вигляд:

$$h_t = tg(w_{xh} \cdot x_t + w_{hh} \cdot h_{t-1} + b_h) \ (2.1)$$

$$y_t = w_{hy} \cdot h_t + b_y \ (2.2)$$

$$o_t = \sigma(y_t) \ (2.3)$$

де:

- x_t - вхідні дані для тимчасового кроку t;

- h_t - прихований стан для тимчасового кроку t;
- y_t - вихідні дані для тимчасового кроку t;
- w - матриці ваг;
- b - вектори зсуву (bias).
- tg - активаційна функція, яка є гіперболічною тангенсною активаційною функцією.
- $\sigma()$ - сигмоїдна активаційна функцію.
- o_t - кінцевий вихід даних.

Виходи o_t, отримані за допомогою сигмоїдної активаційної функції, містять ймовірності між 0 і 1.

Класифікація здійснюється шляхом групування ймовірностей, тобто:

- Клас 1 - це ймовірності між 0 і 0.5.
- Клас 2 - це ймовірності між 0.5 і 1.

Проста модель довготривалої та короткочасної пам'яті (LSTM: Long-short term memory model)

Мережі довготривалої короткочасної пам'яті – це особливий вид RNN, здатний навчатися довготривалим залежностям. LSTM спеціально розроблені для того, щоб уникнути проблеми довготривалої залежності. Запам'ятовування інформації на тривалий час - це практично їхня поведінка за замовчуванням, а не те, чого вони намагаються навчитися [13].

Ми знаємо, що рекурентні нейронні мережі мають вигляд ланцюжка повторюваних модулів нейронної мережі. У стандартних RNN цей повторюваний модуль матиме дуже просту структуру, наприклад, один шар з функцією tg. LSTM ж у свою чергу має цілих 4 шари, котрі зв'язуються між собою у досить цікавий спосіб.

Першим кроком у LSTM, який показано на рис. 1, є прийняття рішення про те, яку інформацію ми збираємося викинути зі стану кожної комірки нейронної мережі. Це рішення приймається сигмоїдним шаром, який називається "шаром забуття". Він опирається на h_{t-1} і x_t, і виводить число від 0 до 1 для кожного

числа у стані комірки C_{t-1}, при цьому 1 означає "повністю зберегти значення", тоді як 0 означає "повністю позбутися значення": [13]

$$f_t = \sigma\left(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f\right) \quad (2.4)$$

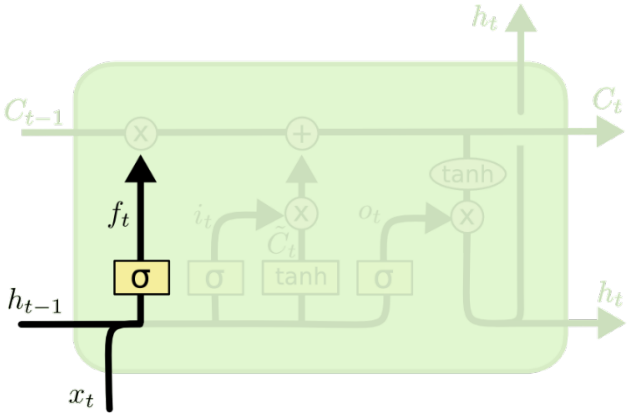

Рис. 2.1. Перший крок моделі LSTM в чарунці C_{t-1}

Джерело: [13]

Наступний крок - вирішити, яку нову інформацію ми збираємося зберігати у стані клітини. Цей процес складається з двох частин. По-перше, сигмоїдний шар вирішує, які значення ми будемо оновлювати, а далі, шар з функцією тангенса створює вектор нових значень-кандидатів C_{t-1} які можна додати до стану. На наступному кроці, що зображено на рис. 2.2 ми об'єднаємо ці два шари, щоб створити оновлення стану чарунки.

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i) \quad (2.5)$$

$$\hat{C}_t = tg(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c) \quad (2.6)$$

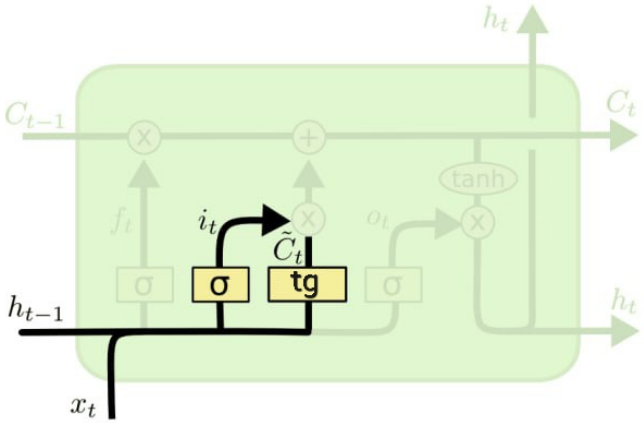

Рис. 2.2. Другий крок моделі LSTM в чарунці C_{t-1}

Джерело: [13]

Тепер на наступному кроці, рис. 2.3, ми переходимо з чарунки C_{t-1} до C_t.

Для цього помножено значення старої чарунки на f_t, відкидаючи певні значення з першого кроку, та додаємо до цього множника $i_t * \hat{C}_t$: [13].

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \ (2.7)$$

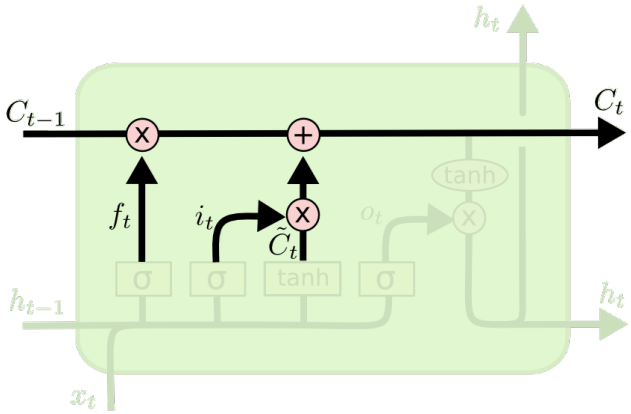

Рис. 2.3. Третій крок моделі LSTM в чарунці C_{t-1}

Джерело: Understanding LSTM Networks [13].

Насамкінець, нам потрібно вирішити, що будемо виводити. Результат базуватиметься на значеннях нашої чарунки, але буде відфільтрованою версією. Спочатку ми запускаємо сигмоїдний шар, який вирішує, які частини значень чарунки ми будемо виводити, потім пропускаємо значення комірки через tg (щоб відфільрувати значення між -1 і 1) і множимо його на вихід сигмоїдного шару, щоб вивести тільки ті частини, які ми вирішили – це видно на рис. 2.4 [13].

$$o_t = \sigma(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o) \ (2.8)$$

$$h_t = o_t * tg(C_t) \ (2.9)$$

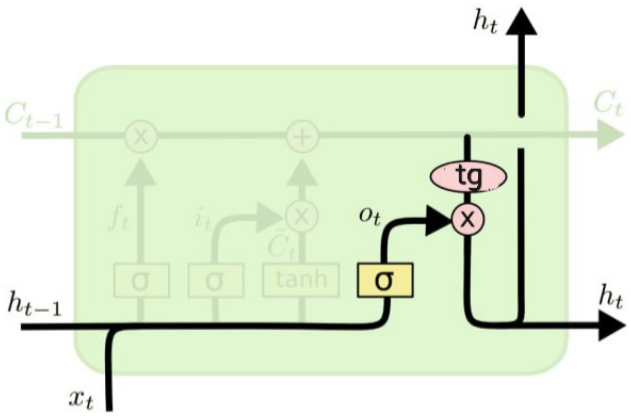

Рис. 2.4. Останній крок моделі LSTM

Джерело: [13]

Двонаправлена модель довготривалої та короткочасної пам'яті (BiLSTM: Bidirectional long-short term memory model)

BiLSTM підвищує ефективність звичайної LSTM, обробляючи вхідні дані за допомогою прямого LSTM і зворотного LSTM. Таким чином вона складається з двох окремих LSTM-мереж: одна обробляє послідовність у прямому напрямку (від початку до кінця), а інша — у зворотному, що можна побачити на рис. 2.5. Результати обох мереж об'єднуються. Це дозволяє моделі при прийнятті рішення для поточного елемента враховувати контекст як з "минулих", так і з "майбутніх" елементів. У нашому випадку, хоч дані й не є послідовними, такий підхід експериментально застосовується для того, щоб кожна ознака була проаналізована в контексті всіх інших.

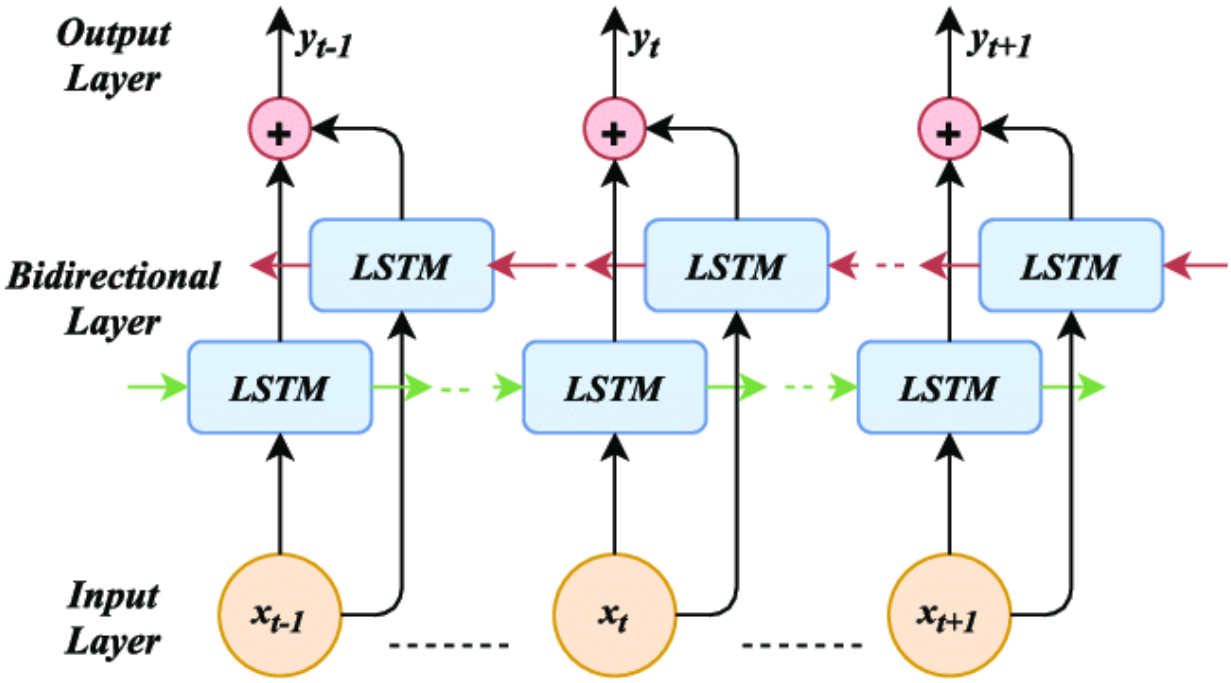

Рис. 2.5. Схематичне представлення BiLSTM для послідовних даних

Джерело: [12]

CNN (згорткова нейронна мережа)

Згорткові нейронні мережі (Convolutional Neural Networks, CNN) — це спеціалізований клас нейронних мереж, розроблений для обробки та аналізу даних, що мають сіткову структуру, таких як зображення та відео. Їхня архітектура натхненна біологічними процесами зорової кори головного мозку

тварин, відповідно даний тип нейронних мереж зазвичай для роботи зі зображеннями.

Основна ідея CNN полягає у виявленні локальних ознак на вхідних даних за допомогою спеціальних фільтрів (ядер) і поступовому узагальненні цих ознак у більш складні патерни на глибших шарах мережі.

2.3. Опис гібридної моделі A-BiLSTM-CNN-Net

A-BiLSTM-CNN-Net модель є складною гібридною нейронною мережею, розробленою для вирішення завдання бінарної класифікації — а саме прогнозування відтоку клієнтів телекомунікаційної компанії. Її ключова особливість полягає в поєднанні кількох передових архітектур глибокого навчання для обробки даних, що дозволяє максимально ефективно виявляти складні залежності між різними характеристиками клієнтів.

Перш ніж дані потрапляють до моделі, вони проходять кілька важливих етапів підготовки, що є критично важливим для якості прогнозу. На початку відбувається етап інженерії ознак, на якому створюються нові, більш інформативні ознаки на основі існуючих. Це допомагає моделі краще зрозуміти поведінку клієнтів. Далі ми почергово обробляємо категорійні ознаки, які перетворюються на числові індекси, та числові ознаки, які масштабуються, щобмати однаковий діапазон значень. Це необхідно, щоб навчання нейронної мережі відбувалося стабільно без екстремальних відхилень.

До того ж набори даних для прогнозування відтоку часто є незбалансованими – клієнтів, що не пішли, зазвичай значно більше, ніж тих, що вдалося утримати. Тому щоб модель не ігнорувала менший клас, а саме клієнтів, що пішли, використовується техніка ADASYN. Вона інтелектуально генерує нові, синтетичні приклади клієнтів, схильних до відтоку, роблячи навчальну вибірку більш збалансованою.

Модель має складну гібридну архітектуру, де кожен блок виконує свою унікальну функцію. Дані проходять через ці блоки послідовно: вхідний гібридний шар (Input Hybrid Layer) – шар вбудовувань (Embedding Layer) – шар

багато-головкової уваги (Multi-Head Self-Attention) – двонаправлений LSTM (Bidirectional LSTM) – 1D-CNN – Classifier Head. Classifier Head.

Вхідний гібридний шар (Input Hybrid Layer) дозволяє обробляти категоріальні та числові ознаки. Так для кожної категорійної ознаки створюється окремий вхід, а числові ознаки подаються всі разом через єдиний вхід.

Категорійні дані не мають внутрішнього числового порядку. Пряме їх кодування (наприклад, 0, 1, 2) може ввести модель в оману, змусивши її думати, що категорія "2" більша за "1". Альтернатива — One-Hot Encoding — створює дуже великі та розріджені вектори, що неефективно. [25]

Шар Embedding вирішує цю проблему. По суті, це таблиця пошуку, параметри якої навчаються. Кожній унікальній категорії зіставляється щільний вектор фіксованої розмірності. Під час тренування моделі ці вектори коригуються за допомогою зворотного поширення помилки. В результаті категорії, які мають схожий вплив на цільову змінну, будуть розташовані ближче одна до одної у цьому новому векторному просторі, як це представлено на рис. 2.6. Це дозволяє моделі вловлювати "семантичну" близькість між категоріями. Наприклад, модель може вивчити, що місячна підписка семантично ближче до піврічної підписка, ніж до підписки на рік з точки зору ризику відтоку.

Щодо опису Multi-Head Self-Attention, то загалом це механізм, що лежить в основі знаменитих моделей-трансформерів, дозволяє вхідним даним "взаємодіяти" між собою. Замість того, щоб розглядати кожну ознаку ізольовано, self-attention дозволяє моделі визначити, на які інші ознаки слід звернути більше "уваги" при обробці поточної. [2]

Це реалізується через три вектори, що навчаються, для кожної ознаки: Query (Запит), Key (Ключ) та Value (Значення). Query – представляє поточну ознаку, яка "шукає" зв'язки; Key – представляє всі інші ознаки, з якими можна встановити зв'язок; Value – містить фактичну інформацію про ці інші ознаки. Модель обчислює "оцінку уваги", порівнюючи Query поточної ознаки з Key кожної іншої. Чим вища оцінка, тим релевантнішою є інша ознака. А Multi-Head означає, що цей процес виконується паралельно кілька разів. Кожна "голова" (head) може

навчитися знаходити різні типи залежностей. Результати всіх голів об'єднуються, що дає багате, зважене уявлення про взаємозв'язки між ознаками.

Classifier Head – це фінальний блок моделі, який приймає оброблений вектор ознак і робить остаточний прогноз.

Для надійної оцінки якості моделі використовується перехресна валідація. Набір даних ділиться на 10 частин. Модель навчається 10 разів, щоразу використовуючи 9 частин для навчання та 1 частину для тестування. Фінальний результат — це усереднене значення по всіх 10 експериментах. Для навчання використовується ефективний оптимізатор Adam – функція втрат Binary Cross-Entropy ідеально підходить для задач бінарної класифікації, вимірюючи розбіжність між прогнозом моделі та реальним результатом. [25]

Також застосовані декілька механізмів контролю: EarlyStopping та ReduceLROnPlateau. Це значить, що навчання автоматично припиняється, якщо якість моделі на валідаційних даних перестає покращуватися, що запобігає перенавчанню; а другий механізм відповідає за те, щоб швидкість навчання динамічно зменшувалась, якщо модель досягає "плато", що дозволяє їй більш точно налаштувати ваги.

Окрім загальної точності (accuracy), для всебічної оцінки моделі використовуються AUC (Area Under the Curve), precision, recall, F1-Score. Це особливо важливо для незбалансованих даних, де точність може бути високою, але оманливою. **Precision** показує точність, яка частка клієнтів, спрогнозованих як такі, що підуть, дійсно пішла. **Recall** вимірює повноту, тобто яку частку клієнтів, що реально пішли, модель змогла правильно ідентифікувати, і він відображає здатність моделі "не пропустити" клієнтів із групи ризику. **F1-Score** є гармонійним середнім між та точністю та повнотою, надаючи єдину збалансовану метрику. **AUC** оцінює загальну здатність моделі розрізняти між класами (відтік / не відтік) незалежно від обраного порогу класифікації. [26]

Таким чином, архітектура A-BiLSTM-CNN-Net, посилена ретельною попередньою обробкою даних, комплексною стратегією валідації, і є гнучким інструментом для прогнозування відтоку. Вона поєднує переваги різних підходів

глибокого навчання для виявлення складних, нелінійних та контекстуальних залежностей у даних, що дозволяє припустити її високу ефективності та надати бізнесу цінну інформацію щодо утримання клієнтів

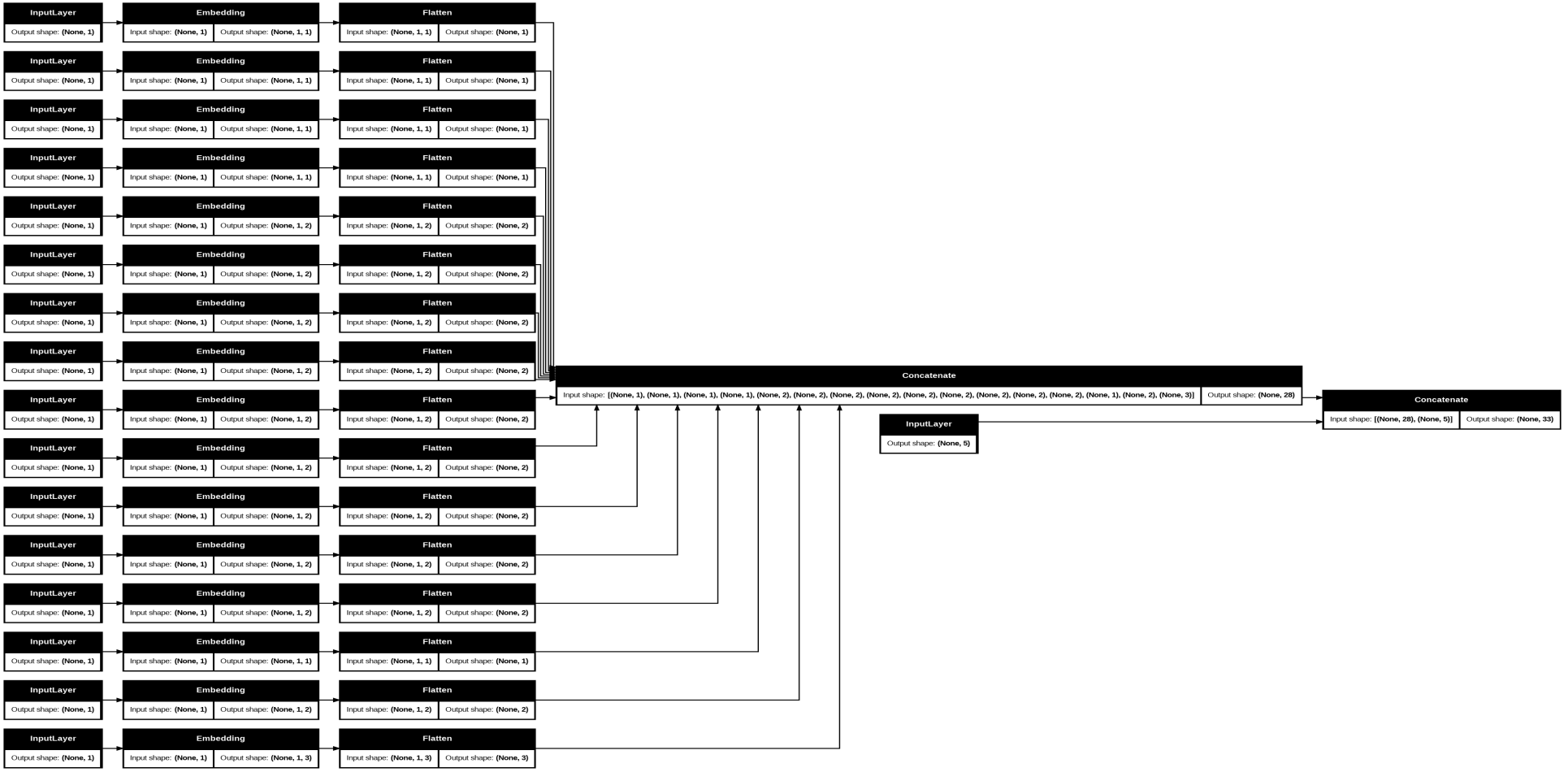

Рис. 2.6. Обробка всіх вхідних даних та створення єдиного вектору ознак

Джерело: розрахунки автора

РОЗДІЛ 3. ПРАКТИЧНЕ ЗАСТОСУВАННЯ ГІБРИДНОЇ МОДЕЛІ ДЛЯ ПРОГНОЗУВАННЯ ВІДТОКУ КЛІЄНТІВ

3.1. Огляд датасету

В цій роботі використані дані семпл даних 7043 клієнтів-користувачів TelCom сфери. [14]

Усі стовпці можна логічно згрупувати в кілька категорій:

1. Демографічна інформація про клієнтів
   - customerID: Унікальний ідентифікатор кожного клієнта.
   - gender: Стать клієнта (Male — чоловік, Female — жінка).
   - SeniorCitizen: Чи є клієнт пенсіонером (1 — так, 0 — ні).
   - Partner: Чи має клієнт партнера (Yes — так, No — ні).
   - Dependents: Чи є клієнт годувальником (наприклад, утримує дітей) (Yes — так, No — ні).
2. Інформація про акаунт клієнта
   - tenure: Кількість місяців, протягом яких клієнт користується послугами компанії.
   - Contract: Тип контракту клієнта (Month-to-month — щомісячний, One year — на один рік, Two year — на два роки).
   - PaperlessBilling: Чи використовує клієнт безпаперове виставлення рахунків (Yes — так, No — ні).
   - PaymentMethod: Спосіб оплати (Electronic check — електронний чек, Mailed check — паперовий чек, Bank transfer (automatic) — банківський переказ, Credit card (automatic) — кредитна картка).
   - MonthlyCharges: Сума щомісячних платежів.
   - TotalCharges: Загальна сума, яку клієнт сплатив за весь час користування послугами.
3. Інформація про підключені послуги
   - PhoneService: Чи підключена послуга телефонного зв'язку (Yes — так, No — ні).

- MultipleLines: Чи підключено кілька телефонних ліній (Yes, No, No phone service).
- InternetService: Тип інтернет-провайдера (DSL, Fiber optic, No — інтернет не підключено).
- OnlineSecurity: Чи підключена послуга онлайн-безпеки (Yes, No, No internet service).
- OnlineBackup: Чи підключена послуга онлайн-резервного копіювання (Yes, No, No internet service).
- DeviceProtection: Чи підключена послуга захисту пристроїв (Yes, No, No internet service).
- TechSupport: Чи підключена послуга технічної підтримки (Yes, No, No internet service).
- StreamingTV: Чи підключена послуга стрімінгового телебачення (Yes, No, No internet service).
- StreamingMovies: Чи підключена послуга стрімінгового кіно (Yes, No, No internet service).

4. Цільова змінна
   - Churn: Показує, чи відмовився клієнт від послуг компанії за останній місяць (Yes — так, No — ні).

Також цей датасет не містить порожніх значень в жодному зі стовпців, а більшість об’єктів подано в текстовому форматі, що видно в табл. 3.1.

Таблиця 3.1

Кількість непорожніх значень для кожного зі стовпців

Стовпець	**К-сть непорожніх значень**	**Dtype**
customerID	7043	object
gender	7043	object
SeniorCitizen	7043	int64
Partner	7043	object
Dependents	7043	Object
tenure	7043	int64
PhoneService	7043	object

Продовження табл. 3.1.

Стовпець	К-сть непорожніх значень	Dtype
MultipleLines	7043	object
InternetService	7043	object
OnlineSecurity	7043	object
OnlineBackup	7043	object
DeviceProtection	7043	object
TechSupport	7043	object
StreamingTV	7043	object
StreamingMovies	7043	object
Contract	7043	object
PaperlessBilling	7043	object
PaymentMethod	7043	object
MonthlyCharges	7043	float64
TotalCharges	7043	object
Churn	7043	object

Як бачимо на рис. 3.1., сильної немає сильної кореляції багатьох ознак, окрім ознак зі значенням, що вказують на відсутність телефонних чи інтернет сервісів.

І також бачимо залежність між TotalCharges та tenure, що є доволі очевидним, адже чим більше клієнт користується сервісами компанії, тим більше коштів він їм приноситиме, що ми й бачимо на рис. 3.2.

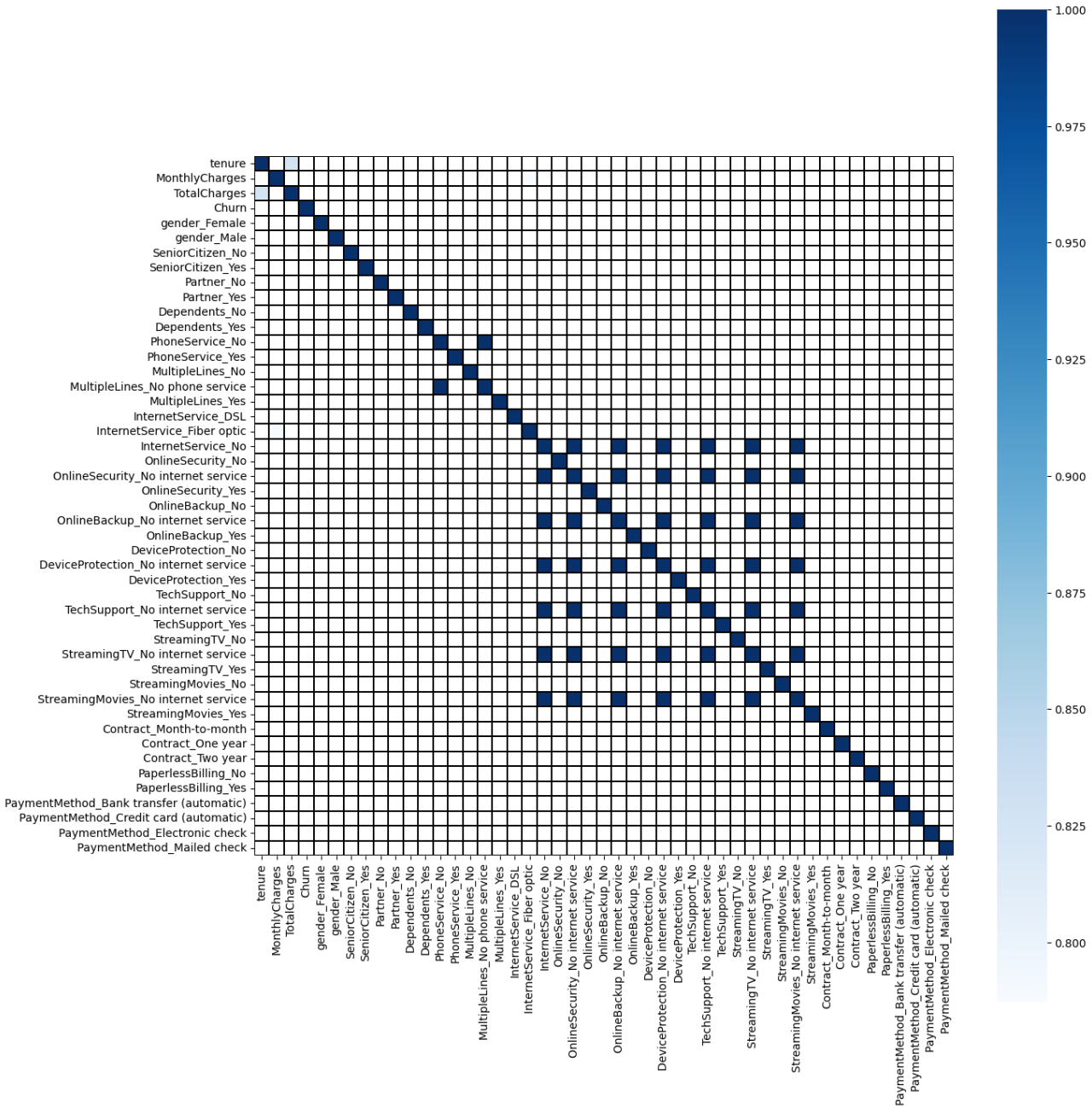

Рис. 3.1. Кореляція між різними ознаками датасету

Джерело: розрахунки автором

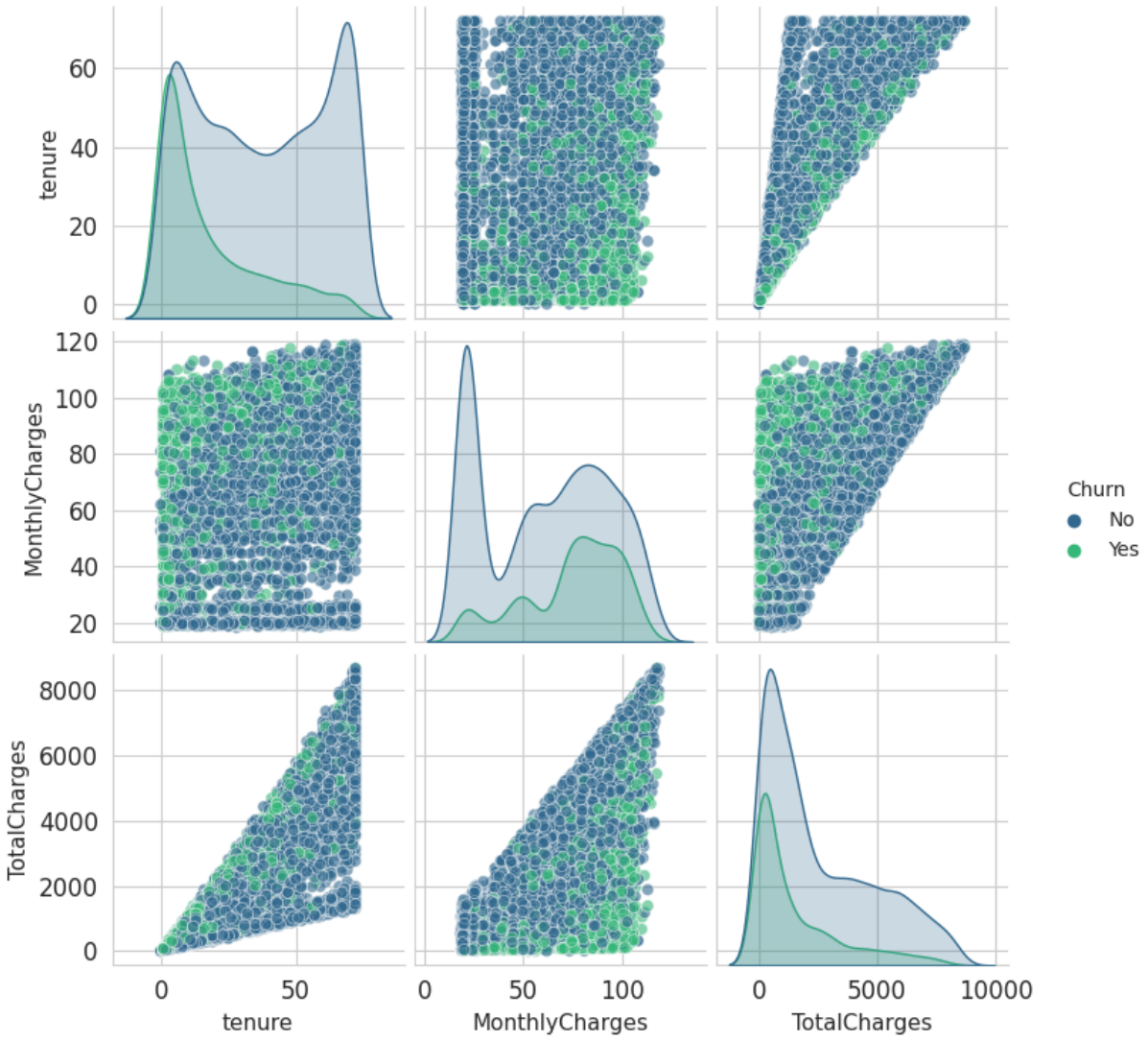

Рис. 3.2. Співвідношення стовпців TotalCharges, MonthlyCharges і Tenure

Джерело: розрахунки автора

Також, з огляду на рис. 3.4., більшість клієнтів є лояльними й лише близько 27% клієнтів відмовилися від послуг за останній період. В свою чергу це означає, що дані є незбалансованими, і лояльних клієнтів значно більше, ніж тих, хто пішов. А на основі рис. 3.3. бачимо, що клієнти, що використовують оптоволоконний інтернет (Fiber optic), йдуть значно частіше, ніж користувачі DSL, що в свою чергу може бути пов'язано з вищою ціною або нестабільністю послуги.

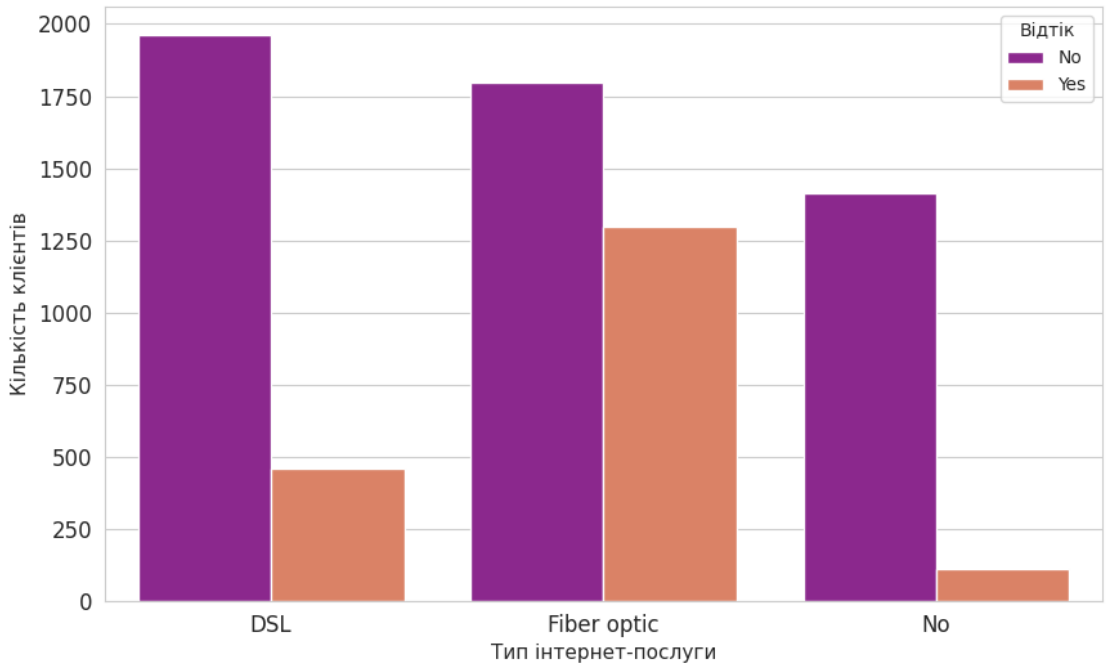

Рис. 3.3. Вплив типу інтернет послуги на відтік

Джерело: розрахунки автором

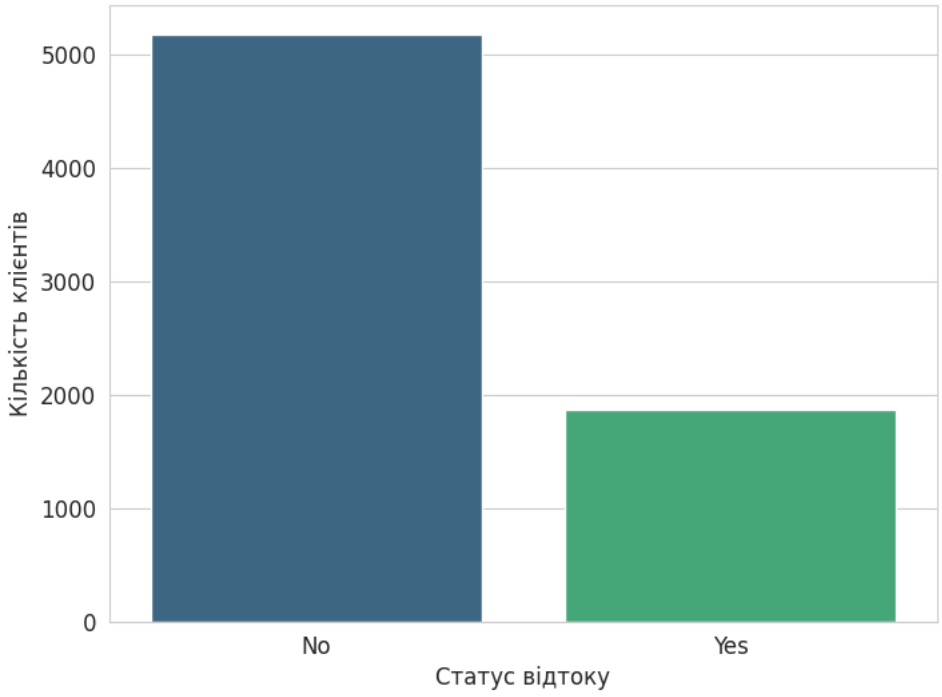

Рис. 3.4. Розподіл стовпця Churn

Джерело: розрахунки автора

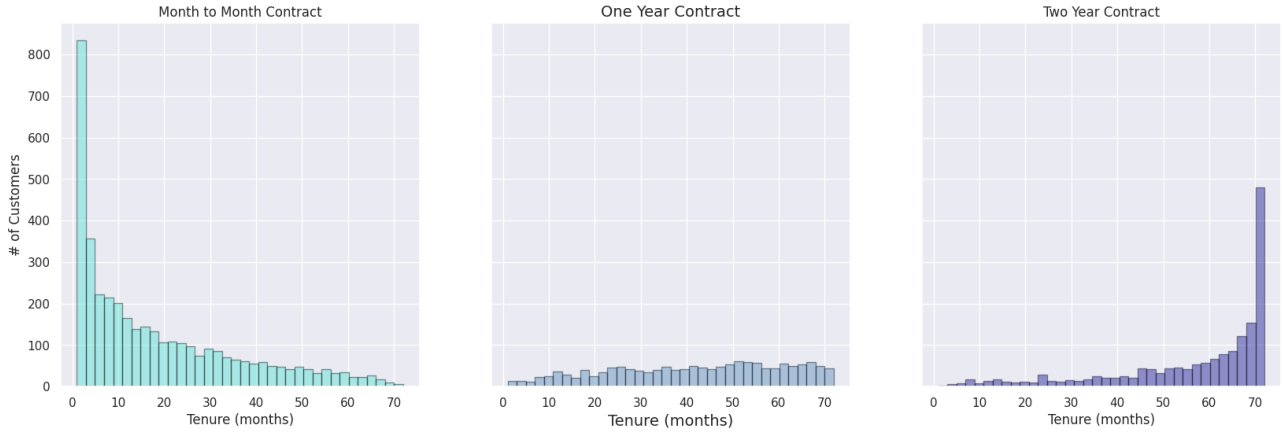

Рис. 3.5. Розподіл тарифних планів за часом утримання клієнтів

Джерело: розрахунки автором

З рис. 3.5. можемо помітити, що клієнти з щомісячним контрактом (Month-to-month) схильні до відтоку набагато більше, ніж клієнти з річними (One year) або дворічними (Two year) контрактами, а пік відтоку припадає на перші місяці користування послугами – в свою чергу клієнти, які залишаються з компанією тривалий час (tenure > 20 місяців), демонструють значно вищу лояльність. Таким чином можна припустити, що довгострокові зобов'язання значно підвищують лояльність й таким чином довше зберігають клієнта в межах власної компанії.

Рис. 3.6. показує, що відтік має негативну кореляцію з tenure (чим довше клієнт з компанією, тим менша ймовірність відтоку) і позитивну з MonthlyCharges (вищі платежі — вища ймовірність відтоку). І також як і на рис. 3.2. бачимо залежність між TotalCharges та tenure, що є доволі очевидним, адже чим більше клієнт користується сервісами компанії, тим більше коштів він їм приноситиме.

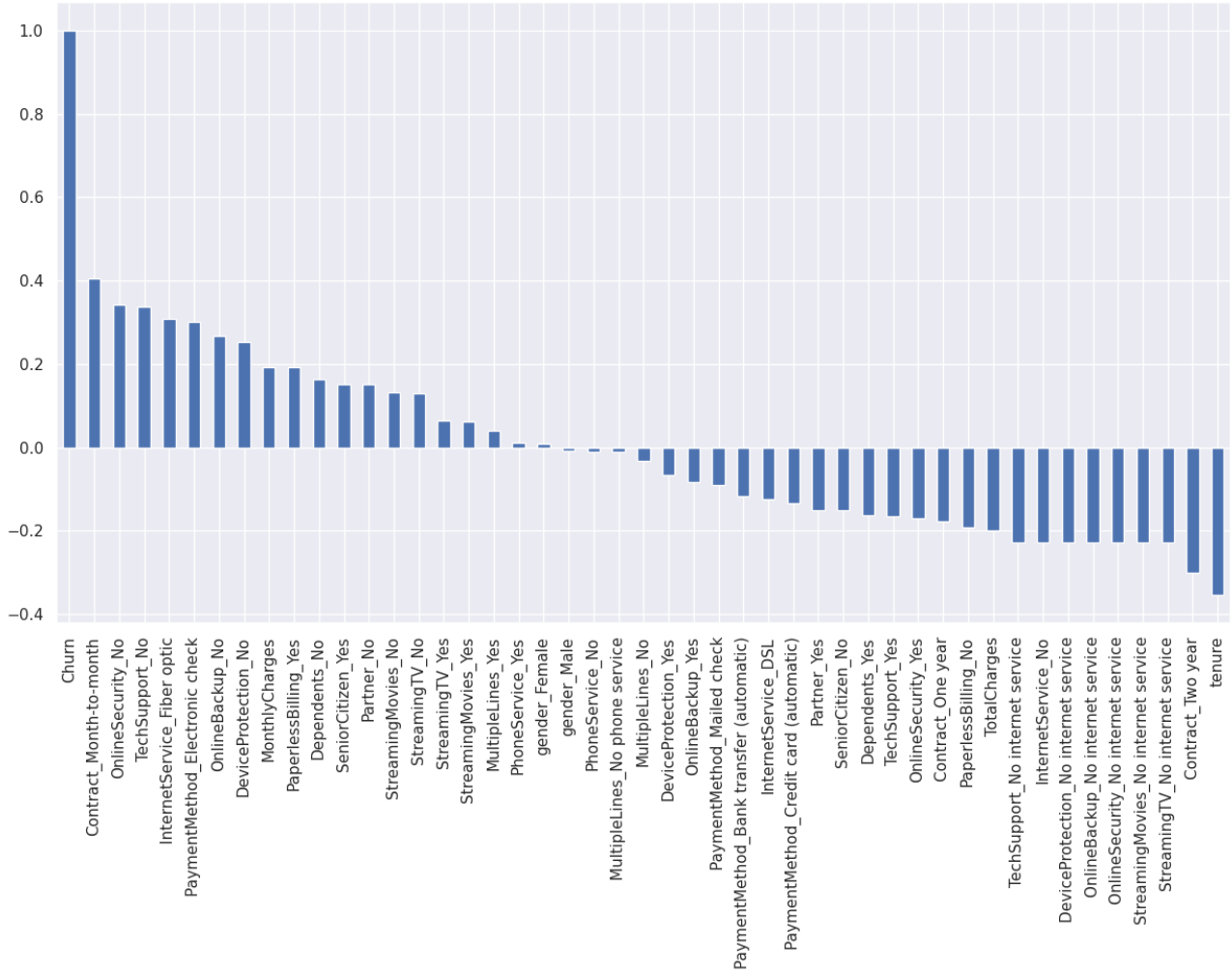

Рис. 3.6. Кореляції всіх атрибутів зі стовпцем Churn

Джерело: розрахунки автором

Отже, якщо спробувати сформувати профіль найбільш схильного до відтоку клієнта, відповідно до цього датасету, я б виділив такі його характерні риси:

- має короткостроковий щомісячний контракт;
- користується послугою оптоволоконного інтернету (Fiber optic);
- є відносно новим клієнтом (користується послугами менше року);
- має високий щомісячний платіж.

Для зниження відтоку компанії варто звернути увагу саме на цю групу клієнтів, можливо, запропонувавши їм вигідніші умови при переході на довгостроковий контракт або проаналізувавши якість послуги оптоволоконного інтернету.

3.2. Підготовка даних та тренування моделі

Спочатку з набору даних видаляється ідентифікатор customerID, оскільки він не містить прогнозної інформації. Згідно з табл. 3.1., стовпець TotalCharges також містить текстові дані, хоча очікувані значення – числові, тому перетворимо його на відповідний числовий формат, а значення, котрі містять невалідні дані, такі як пробіли – таких усього 11 рядків – замінюються на пропуски (NaN). Ці пропуски заповнюються медіанним значенням стовпця. Вибір медіани є обґрунтованим, оскільки вона є більш стійкою до статистичних викидів, на відміну від середнього арифметичного. І, звісно, цільова змінна Churn перетворюється на бінарний формат (1/0) відповідно до вимог задачі класифікації, де 1 позначає, що клієнт припинив користуватися сервісами компанії, а 0 – що клієнт не припиняв користуватися сервісами компанії. [22]

Далі для потенційного підвищення прогнозної здатності моделі створюються дві нові ознаки: tenure_group та monthly_to_total_ratio. Перша – це дискретизована, тобто розбита на п'ять категорійних інтервалів, числова ознака tenure. Дана ознака дозволяє моделі виявляти нелінійні залежності між тривалістю контракту та цільовою змінною. Друга, monthly_to_total_ratio – відношення місячних платежів до загальних. На фінальному етапі підготовки дані уніфікуються. Категорійні ознаки, включно зі створеною tenure_group, перетворюються на числові за допомогою *LabelEncoder*. Усі числові ознаки проходять процедуру стандартизації за допомогою *StandardScaler*. Ця операція приводить кожну ознаку до нульового середнього та одиничної дисперсії, що є важливою передумовою для стабільної та швидкої збіжності алгоритмів оптимізації.

Процес тренування організований таким чином, щоб забезпечити об'єктивну оцінку продуктивності моделі та мінімізувати ризик перенавчання. В якості перехресної валідації застосовується метод *K-Fold* з k=10. Даний підхід передбачає розбиття всього набору даних на 10 частин (фолдів): тобто модель тренується 10 разів, де на кожній ітерації для тренування використовуються 9 фолдів, а для валідації — 1 фолд, що залишився. В кінці фінальні метрики якості,

AUC, усереднюються по всіх ітераціях. Це дозволяє отримати більш стійку та генералізовану оцінку ефективності моделі порівняно з єдиним розбиттям на тренувальний та тестовий набір.

Також, оскільки у вихідних даних спостерігається дисбаланс класів. Для його пом'якшення на кожному етапі тренування застосовується алгоритм ADASYN. Проте важливо зазначити, що ADASYN застосовується виключно до тренувального набору даних усередині кожного фолду, що запобігає витоку даних у валідаційний набір.

3.3. Оцінка якості гібридної моделі

Для оцінки якості гібридної моделі було використано ряд показників, зокрема Precision, Recall, F1-Score та AUC. Результати гібридної моделі порівнювалися з базовими Random Forest, XGBoost, Gradient Boosting, SVM, KNN – які використовуються в якості орієнтиру.

При поверхневому аналізі метрик з табл. 3.2, моделі Random Forest та XGBoost демонструють найвищу збалансовану продуктивність. Зокрема, їхні показники F1-Score для ключового класу "Churn" (0.81 та 0.78 відповідно) значно перевищують результати інших моделей. Гібридна модель, процес тренування якох можна побачити на рис. 3.10., з F1-Score 0.62, на перший погляд, виглядає суттєво слабшою, показуючи ефективність, вищу лише за SVM. Цей низький показник F1-Score у гібридної моделі зумовлений значним дисбалансом між відносно високою повнотою (Recall = 0.71) та низькою точністю.

Таблиця 3.2

Метрики якості моделей

Модель	Категорія	Precision	Recall	F1-Score
Random Forest Classifier	Not Churn	0.92	0.96	0.94
Random Forest Classifier	Churn	0.86	0.77	0.81
Random Forest Classifier	Weighted Avg	0.9	0.9	0.9
XGBoost Classifier	Not Churn	0.91	0.94	0.92
XGBoost Classifier	Churn	0.81	0.75	0.78
XGBoost Classifier	Weighted Avg	0.89	0.89	0.89

Продовження табл. 3.2.

Модель	Категорія	Precision	Recall	F1-Score
Gradient Boosting Classifier	Not Churn	0.86	0.91	0.88
Gradient Boosting Classifier	Churn	0.71	0.58	0.64
Gradient Boosting Classifier	Weighted Avg	0.82	0.82	0.82
Hybrid model	Not Churn	0.88	0.79	0.83
Hybrid model	Churn	0.55	0.71	0.62
Hybrid model	Weighted Avg	0.79	0.77	0.78
Support Vector Machine (SVM)	Not Churn	0.84	0.91	0.87
Support Vector Machine (SVM)	Churn	0.68	0.51	0.58
Support Vector Machine (SVM)	Weighted Avg	0.8	0.81	0.8
Logistic Regression	Not Churn	0.85	0.9	0.87
Logistic Regression	Weighted Avg	0.8	0.8	0.8
Logistic Regression	Churn	0.66	0.55	0.6
K-Nearest Neighbors (KNN)	Not Churn	0.85	0.87	0.86
K-Nearest Neighbors (KNN)	Churn	0.61	0.57	0.59
K-Nearest Neighbors (KNN)	Weighted Avg	0.78	0.79	0.79

Однак, аналіз інтегральної метрики AUC, що відображає загальну роздільну здатність моделей незалежно від порогу класифікації, представляє картину під іншим кутом. Порівнюючи площу під ROC-кривими на рис. 3.7, 3.8 та 3.9, ми бачимо, що гібридна модель демонструє найкращий результат (**AUC = 0.83**), випереджаючи Random Forest та XGBoost (обидві з **AUC ≈ 0.82**). Високий показник AUC (0.83) свідчить про те, що фундаментальна здатність гібридної моделі правильно ранжувати клієнтів за ступенем ризику є найкращою серед усіх розглянутих. Це означає, що вона найімовірніше присвоїть вищу оцінку ризику клієнту, який дійсно піде, ніж клієнту, який залишиться.

Це робить гібридну модель чудовим інструментом для задач, де важлива пріоритизація, наприклад, для формування списку клієнтів для утримання.

Гібридна модель має показник AUC на рівні 0.8322, що вказує на високу роздільну здатність моделі, тобто її здатність правильно ранжувати клієнтів за ймовірністю відтоку. Модель значно краща за випадкове вгадування, при якому AUC = 0.5. Низьке стандартне відхилення (+/- 0.0106 для AUC) свідчить про те,

що модель працює стабільно на різних частинах даних, а результати не є випадковими і добре узагальнюються.

І хоч показник точності Accuracy ~76% є непоганим, але AUC є більш надійною метрикою для задач з потенційним дисбалансом класів (що в нашому випадку беззаперечно спостерігається), оскільки точність може бути високою, навіть якщо модель погано прогнозує менший клас.

Модель досягає оптимальних або близьких до оптимальних результатів досить швидко. У більшості фолдів найкраща val_loss (втрати на валідації) досягається в межах перших 20-30 епох.

Важливо, що val_loss не починає стрімко зростати, що свідчить про ефективність механізмів регуляризації (шари Dropout) та ранньої зупинки. [20] Тобто модель добре узагальнює знання, а не просто "запам'ятовує" тренувальні дані.

За допомогою ADASYN перед початком тренування на кожному фолді вибірка була майже ідеально збалансована. Саме цей крок змушує модель приділяти однакову увагу обом класам і є ключовим фактором досягнення високого показника AUC. Без цього модель, ймовірно, ігнорувала б менший клас (Churn=1).

Результати для кожного з 10 фолдів є досить близькими: AUC варіюється в діапазоні приблизно від 0.816 (Фолд 8) до 0.852 (Фолд 4).

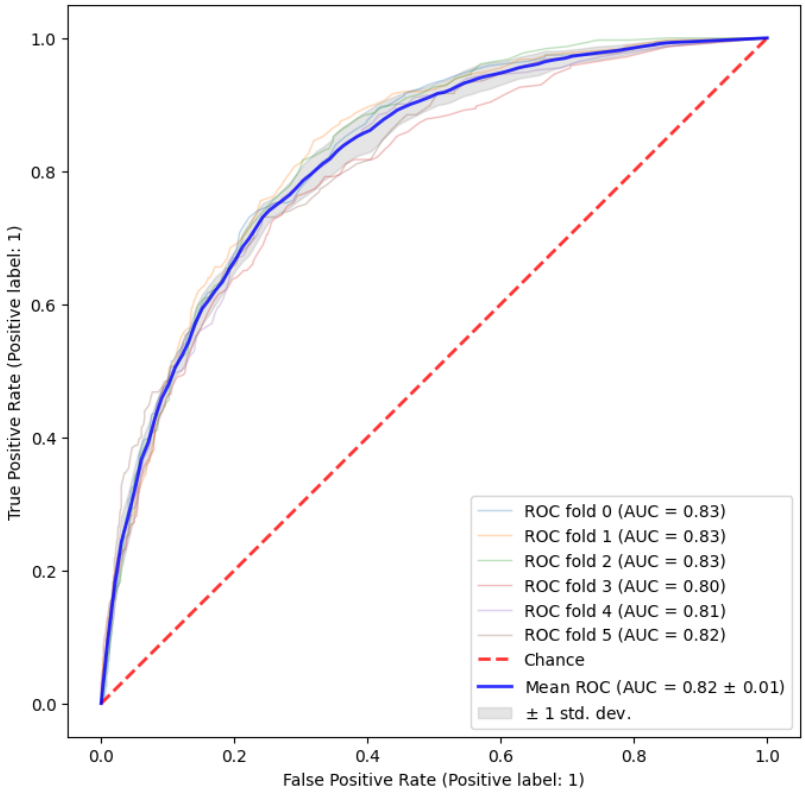

Рис. 3.7. ROC крива для моделі Random Forest

Джерело: розрахунки автора

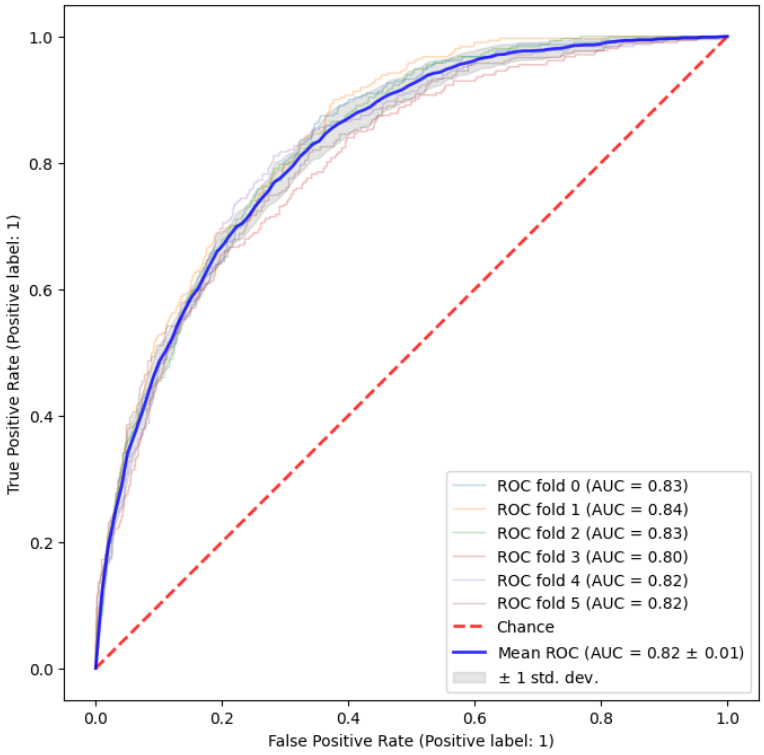

Рис. 3.8. ROC крива для моделі XGBoost

Джерело: розрахунки автора

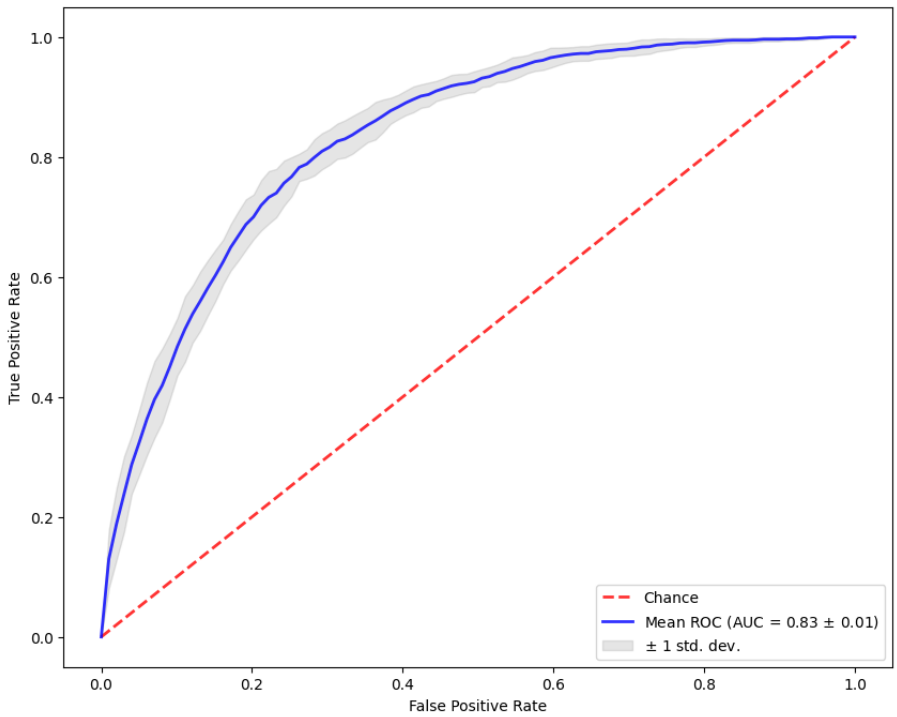

Рис. 3.9. ROC крива для гібридної моделі

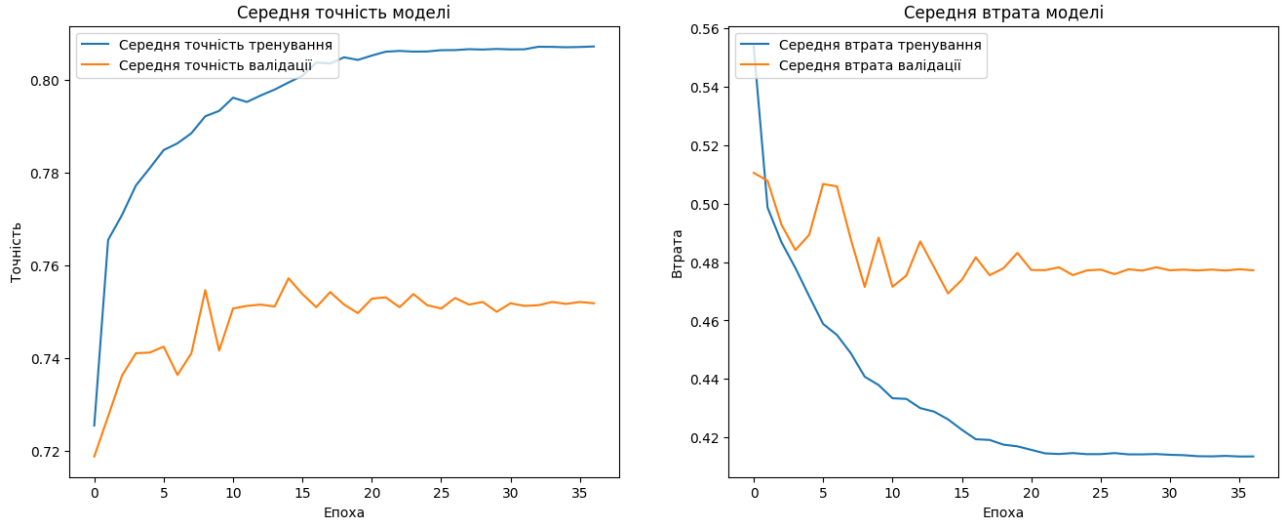

Рис. 3.10. Динаміка тренувань моделі

Джерелло: розрахунки автора

Отже, хоч і стандартні моделі, такі як Random Forest та XGBoost, добре працюють при стандартному порозі класифікації й без особливих додаткових налаштувань, наведена гібридна модель при порівнянні метрики AUC показала себе найкраще серед інших моделей, що свідчить про її високий потенціал до ранжування. Таким чином гібридна модель при додатковому налаштуванні ряду параметрів, має потенціал обійти інші моделі не тільки за метрикою AUC, але й за усіма іншими метриками також.

ВИСНОВКИ

Як результат даної роботи, було детально проаналізовано проблему відтоку клієнтів, її економічний вплив на бізнес та було розглянуто актуальність застосування нейронних мереж для прогнозування відтоку клієнтів.

Сформовано профіль найбільш схильного до відтоку клієнту, згідно з розглянутим датасетом – клієнт, що не брав на себе тривалих зобов'язань, тобто підписки є короткостроковими, користується послугами менше року та має високий щомісячний платіж.

Маючи відповідні висновки, було застосовано гібридну модель A-BiLSTM-CNN-Net для прогнозування відтоку користувачів телекомунікаційної галузі, після чого проведено порівняння з базовими моделями такі як Random Forest, XGBoost, SVM, Linear Classifier.

За результатами аналізу та порівняння роботи моделей, можна зробити висновок, що гібридна модель є досить перспективною для подальшого розвитку щодо прогнозування відтоку клієнтів. І хоч Random Forest та XGBoost є врівноваженими моделями [18], які добре працюють при стандартному порозі класифікації й без великих зусиль на додаткові налаштування, гібридна модель, у свою чергу, має найвищий потенціал до ранжування й показала себе найкраще серед інших моделей при порівнянні метрики AUC. При додаткових пошуках порогу та налаштуванні, дана модель має потенціал обійти інші моделі вже за усіма іншими метриками також.

Поєднання MultiHeadAttention, BiLSTM та CNN дозволяє моделі одночасно виявляти глобальні залежності, довгострокові контекстуальні зв'язки та локальні патерни в даних клієнта, що недоступно для простіших моделей. Використання техніки ADASYN дозволяє значно покращити якість прогнозу для меншого класу (клієнтів, що йдуть), що є критично важливим для бізнес-задач. Архітектура є модульною, що дозволяє легко адаптувати або змінювати її компоненти для вирішення інших схожих завдань.

Таким чином, дана гібридна модель може бути розгорнута як мікросервіс (наприклад, за допомогою TensorFlow Serving або FastAPI) для надання прогнозів

у реальному часі, інтегруючись у CRM-системи компанії для проактивної роботи з клієнтами.

В якості рекомендації бізнесу, зокрема відповідно до даних з наявного датасету, компаніям варто звертати увагу групу клієнтів, що тільки-но почала користуватися сервісами компанії, можливо, запропонувавши їм вигідніші умови при переході на довгостроковий контракт – таким чином уклавши довгострокове зобов’язання, яке користувач з меншою ймовірністю буде розривати. А впровадивши в робочий процес мікросервіс для надання прогнозів у реальному часі, компанія може автоматизувати відбір таких клієнтів й намагатися активніше працювати з ними й пропонувати вигідні пропозиції.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ranktracker. 55+ Must-Know Customer Retention Statistics to Boost Your Business in 2024. URL: https://www.ranktracker.com/blog/55-must-know-customer-retention-statistics/
2. S. Bhattacharjee, U.Thukral. Early Churn Prediction from large scale user-product interaction time series. URL: https://arxiv.org/pdf/2309.14390
3. Omer Faruk Seymen, Abdulkadir Hiziroglu, Onur Dogan, Orhan Er. Customer Churn Prediction Using Ordinary Artificial Neural Network and Convolutional Neural Network Algorithms: A Comparative Performance Assessment. URL: https://www.researchgate.net/publication/360529236_Customer_Churn_Prediction_Using_Ordinary_Artificial_Neural_Network_and_Convolutional_Neural_Network_Algorithms_A_Comparative_Performance_Assessment
4. Payproglobal: Що таке коефіцієнт відтоку клієнтів SaaS? URL: https://payproglobal.com/uk/відповіді/що-таке-коефіцієнт-відтоку-saas/
5. Sulaiman Olaniyi Abdulsalam, Micheal Arowolo, Yakub Kayode Saheed, Jesutofunmi Onaope Afolayan: Customer Churn Prediction in Telecommunication Industry Using Classification and Regression Trees and Artificial Neural Network Algorithms. URL: https://www.researchgate.net/publication/361838640_Customer_Churn_Prediction_in_Telecommunication_Industry_Using_Classification_and_Regression_Trees_and_Artificial_Neural_Network_Algorithms
6. Luca Longo. A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners: a literature review and a research agenda. URL: https://www.researchgate.net/publication/380642749_A_Review_on_M

achine_Learning_Methods_for_Customer_Churn_Prediction_and_Recommendations_for_Business_Practitioners

7. What is an average churn rate? Here's how to figure it out | Stripe, https://stripe.com/ae/resources/more/what-is-an-average-churn-rate-here-is-how-to-figure-it-out
8. InMoment: Customer Churn: How to Predict & Prevent. URL: https://inmoment.com/blog/customer-churn/
9. Mailchimp: Customer Churn: Definition and How to Reduce It. URL: https://mailchimp.com/resources/customer-churn/
10. ScaleMath What is a Good Monthly Churn Rate? Industry Benchmarks. URL: https://scalemath.com/blog/what-is-a-good-monthly-churn-rate/
11. Abinash Mishra, U. Srinivasulu Reddy: A Comparative Study of Customer Churn Prediction in Telecom Industry Using Ensemble Based Classifiers. URL: https://www.researchgate.net/publication/325419986_A_comparative_study_of_customer_churn_prediction_in_telecom_industry_using_ensemble_based_classifiers
12. Multimodal Sensing Devices, Augustine O. Nwajana A Deep Learning Approach for Human Activities Recognition From. URL: https://www.researchgate.net/figure/Bidirectional-LSTM-model-showing-the-input-and-output-layers-The-red-arrows-represent_fig3_344554659
13. Christopher Olah: Understanding LSTM Networks. 2015. URL: https://colah.github.io/posts/2015-08-Understanding-LSTMs/
14. Kaggle: Telco Customer Churn Dataset. URL: https://www.kaggle.com/datasets/blastchar/telco-customer-churn
15. Abadi M., Agarwal A., Barham P., Brevdo E., Chen Z., Citro C., Corrado G. S., Davis A., Dean J. and others TensorFlow: A System for Large-Scale Machine Learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*. 2016. P. 264-

283. URL: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

16. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Müller A. and others Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2012. P. 2825-2830. URL: https://arxiv.org/abs/1201.0490
17. Dalia shihab Ahmed, Rasha shaker ibrahim Al-badri, Firas Ali Hashim, Nadia Mahmood Hussien: Keras deep learning package in python: a review. 2023.P. 24-29. URL: https://www.researchgate.net/publication/374023047_KERAS_DEEP_LEARNING_PACKAGE_IN_PYTHON_A_REVIEW
18. Liashenko, O., Kravets, T. & Kostovetskyi, Y. Machine Learning and Data Balancing Methods for Bankruptcy Prediction, Ekonomika, 2023.
19. Вербівський, Д. С., Карплюк, С. О., Фонарюк, О. В., Сікора, Я. Б. Бустінг і беггінг як методи формування ансамблей моделей. *Abstracts of the 7th International scientific and practical conference*. P. 163-169. URL: http://eprints.zu.edu.ua/32426/
20. Yathish V.: Loss Functions and Their Use In Neural Networks. Towards Data Science. 2022. URL: https://towardsdatascience.com/loss-functions-and-their-use-in-neural-networks-a470e703f1e9
21. The transmitted: Що таке багатошаровий персептрон (Multilayer Perceptron, MLP) у машинному навчанні?. 2023. URL: https://thetransmitted.com/adlucem/shho-take-mlp-u-mashynnomu-navchanni/
22. Mrs. D. Hima Bindu, Mr. P. Ratna Rohith, Mr. P. Hemanth Nani, Ms. K. Niharika, Mr. T. Vinay: Customer Churn Prediction Using Custom Recurrent Neural Network Model. URL: https://ijirt.org/publishedpaper/IJIRT174604_PAPER.pdf
23. Amjad Hudaib, Reham Dannoun, Osama Harfoushi, Ruba Obiedat, Hossam Faris: Hybrid Data Mining Models for Predicting Customer

Churn - Scientific Research Publishing, https://www.scirp.org/journal/paperinformation?paperid=55467

24. GeeksforGeeks: RNN vs LSTM vs GRU vs Transformers. 2025 URL: https://www.geeksforgeeks.org/rnn-vs-lstm-vs-gru-vs-transformers/
25. upGrad: Neural Network Architecture: Types, Components & Key Algorithms. 2025 URL: https://www.upgrad.com/blog/neural-network-architecture-components-algorithms/
26. Jakub Czakon: F1 Score vs ROC AUC vs Accuracy vs PR AUC: Which Evaluation Metric Should You Choose? URL: https://neptune.ai/blog/f1-score-accuracy-roc-auc-pr-auc

ДОДАТКИ

Додаток А

Важливі блоки програмного коду

```
# seventh iteration
import pandas as pd
import numpy as np
import tensorflow as tf
from tensorflow.keras.models import Model
from tensorflow.keras.layers import (Input, Dense, Embedding, Concatenate, Flatten, Dropout,
                                     BatchNormalization, Reshape, Bidirectional, LSTM,
                                     Conv1D, GlobalMaxPooling1D, MultiHeadAttention)
from tensorflow.keras.callbacks import EarlyStopping, ReduceLROnPlateau
from sklearn.model_selection import KFold
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.metrics import classification_report, confusion_matrix, roc_auc_score, accuracy_score
from imblearn.over_sampling import ADASYN
import matplotlib.pyplot as plt
import seaborn as sns


try:
    df=pd.read_csv('/kaggle/input/telco-customer-churn/WA_Fn-UseC_-Telco-Customer-Churn.csv')
except FileNotFoundError:
    print("Помилка: 'WA_Fn-UseC_-Telco-Customer-Churn.csv' не знайдено.")
    print("Будь ласка, переконайтеся, що файл набору даних знаходиться в тій же директорії, що і скрипт.")
    exit()
```

```
df = df.drop('customerID', axis=1)

df['TotalCharges'] = pd.to_numeric(df['TotalCharges'], errors='coerce')
df['TotalCharges'].fillna(df['TotalCharges'].median(), inplace=True)

df['Churn'] = df['Churn'].apply(lambda x: 1 if x == 'Yes' else 0)

print("Проведення feature engineering...")
bins = [0, 12, 24, 48, 60, 72]
labels = ['0-1 Year', '1-2 Years', '2-4 Years', '4-5 Years', '5+ Years']
df['tenure_group'] = pd.cut(df['tenure'], bins=bins, labels=labels, right=False)
df['monthly_to_total_ratio'] = df['MonthlyCharges'] / (df['TotalCharges'] + 1)

categorical_cols = [col for col in df.columns if df[col].dtype == 'object' or
pd.api.types.is_categorical_dtype(df[col].dtype)]
numerical_cols = [col for col in df.columns if df[col].dtype in ['int64', 'float64'] and col
!= 'Churn']

print(f"Категорійні колонки: {categorical_cols}")
print(f"Числові колонки: {numerical_cols}")

label_encoders = {}
for col in categorical_cols:
    le = LabelEncoder()
    df[col] = le.fit_transform(df[col])
    label_encoders[col] = le

scaler = StandardScaler()
df[numerical_cols] = scaler.fit_transform(df[numerical_cols])
```

```
X = df.drop('Churn', axis=1)
y = df['Churn']

feature_cols = X.columns.tolist()
categorical_cols = [col for col in feature_cols if col in categorical_cols]
numerical_cols = [col for col in feature_cols if col not in categorical_cols]

def build_hybrid_model():
    # Model Inputs
    inputs_list = []
    embeddings = []

    for col in categorical_cols:
        num_unique_values = X[col].nunique()
        embedding_dim = int(min(np.ceil(num_unique_values / 2), 50))

        input_cat = Input(shape=(1,), name=f'input_{col}')
        inputs_list.append(input_cat)

        embedding = Embedding(input_dim=num_unique_values,
output_dim=embedding_dim, name=f'embedding_{col}')(input_cat)
        embedding = Flatten(name=f'flatten_{col}')(embedding)
        embeddings.append(embedding)

    input_num = Input(shape=(len(numerical_cols),), name='input_numerical')
    inputs_list.append(input_num)

    concat_embeddings = Concatenate(name='concatenate_embeddings')(embeddings)
    concat_all = Concatenate(name='concatenate_all')([concat_embeddings,
input_num])
```

```
    reshaped_for_sequence = Reshape((1, concat_all.shape[1]))(concat_all)

    attention_dim = 64
    num_heads = 4
    attention_out = MultiHeadAttention(num_heads=num_heads, key_dim=attention_dim)(reshaped_for_sequence, reshaped_for_sequence)

    lstm_out = Bidirectional(LSTM(100, return_sequences=True, recurrent_dropout=0.2))(attention_out)
    cnn_out1 = Conv1D(filters=100, kernel_size=1, padding='same', activation='relu')(lstm_out)
    cnn_out2 = Conv1D(filters=64, kernel_size=1, padding='same', activation='relu')(cnn_out1)
    pool_out = GlobalMaxPooling1D()(cnn_out2)

    x = Dense(128, activation='relu')(pool_out)
    x = Dropout(0.5)(x)
    x = Dense(64, activation='relu')(x)
    x = Dropout(0.4)(x)
    output = Dense(1, activation='sigmoid', name='output')(x)

    model = Model(inputs=inputs_list, outputs=output)

    model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=0.001),
                  loss='binary_crossentropy',
                  metrics=['accuracy', tf.keras.metrics.AUC(name='auc')])

    return model
```

```
n_splits = 10
kfold = KFold(n_splits=n_splits, shuffle=True, random_state=42)

fold_accuracies = []
fold_aucs = []
fold_histories = []

print(f"\nПочаток {n_splits}-кратної перехресної валідації з ADASYN...")

for fold, (train_idx, val_idx) in enumerate(kfold.split(X, y)):
    print(f"--- Фолд {fold+1}/{n_splits} ---")

    X_train, X_val = X.iloc[train_idx], X.iloc[val_idx]
    y_train, y_val = y.iloc[train_idx], y.iloc[val_idx]

    print("Застосування ADASYN до тренувальних даних...")
    adasyn = ADASYN(random_state=42)
    X_train_resampled, y_train_resampled = adasyn.fit_resample(X_train, y_train)
    print(f"Розподіл класів після ADASYN:
\n{pd.Series(y_train_resampled).value_counts()}")

    model = build_hybrid_model()

    X_train_resampled_df = pd.DataFrame(X_train_resampled, columns=X.columns)
    X_train_dict = {f'input_{col}': X_train_resampled_df[col].values for col in
categorical_cols}
    X_train_dict['input_numerical'] = X_train_resampled_df[numerical_cols].values

    # Validation data remains original (not resampled)
    X_val_dict = {f'input_{col}': X_val[col].values for col in categorical_cols}
```

```
    X_val_dict['input_numerical'] = X_val[numerical_cols].values

    reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.2, patience=5, min_lr=0.00001, mode='min')
    early_stopping = EarlyStopping(monitor='val_loss', patience=10, mode='min', restore_best_weights=True)

    history = model.fit(X_train_dict, y_train_resampled,
                        epochs=150,
                        batch_size=64,
                        validation_data=(X_val_dict, y_val),
                        callbacks=[early_stopping, reduce_lr],
                        verbose=1)

    fold_histories.append(history)

    y_pred_proba = model.predict(X_val_dict)
    y_pred = (y_pred_proba > 0.5).astype(int)

    acc = accuracy_score(y_val, y_pred)
    auc = roc_auc_score(y_val, y_pred_proba)

    fold_accuracies.append(acc)
    fold_aucs.append(auc)

    print(f"Результати для фолду {fold+1}: Accuracy = {acc:.4f}, AUC = {auc:.4f}")

print("\n--- Результати перехресної валідації ---")
print(f"Середня точність (Accuracy): {np.mean(fold_accuracies):.4f} (+/- {np.std(fold_accuracies):.4f})")
```

```
print(f"Середній AUC: {np.mean(fold_aucs):.4f} (+/- {np.std(fold_aucs):.4f})")

def plot_avg_history(histories):
    avg_acc = np.mean([h.history['accuracy'] for h in histories], axis=0)
    avg_val_acc = np.mean([h.history['val_accuracy'] for h in histories], axis=0)
    avg_loss = np.mean([h.history['loss'] for h in histories], axis=0)
    avg_val_loss = np.mean([h.history['val_loss'] for h in histories], axis=0)

    fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(16, 6))

    ax1.plot(avg_acc, label='Середня точність тренування')
    ax1.plot(avg_val_acc, label='Середня точність валідації')
    ax1.set_title('Середня точність моделі')
    ax1.set_ylabel('Точність')
    ax1.set_xlabel('Епоха')
    ax1.legend(loc='upper left')

    ax2.plot(avg_loss, label='Середня втрата тренування')
    ax2.plot(avg_val_loss, label='Середня втрата валідації')
    ax2.set_title('Середня втрата моделі')
    ax2.set_ylabel('Втрата')
    ax2.set_xlabel('Епоха')
    ax2.legend(loc='upper left')

    plt.show()

max_epochs = max(len(h.history['accuracy']) for h in fold_histories) if fold_histories
else 0
# Pad shorter histories to allow for averaging
padded_histories = []
```

```
for h in fold_histories:
    history_dict = {}
    for key, values in h.history.items():
        pad_width = max_epochs - len(values)
        padded_values = np.pad(values, (0, pad_width), 'edge')
        history_dict[key] = padded_values
    padded_histories.append({'history': history_dict})
```