

УДК 004.942

DOI: 10.20535/1810-0546.2017.5.110338

Н.В. Кузнецова, П.І. Бідюк*
КПІ ім. Ігоря Сікорського, Київ, Україна

МОДЕЛЮВАННЯ ФІНАНСОВОГО РИЗИКУ В ТЕЛЕКОМУНІКАЦІЙНІЙ СФЕРІ

Background. The telecommunication field in Ukraine is dynamically developing continuously renewing its proposals for the market and consumer requirements. That is why a timely estimation of financial risks and optimization of financial expenses regarding development of new components and possible losses of clients is especially urgent problem today.

Objective. The aim of the paper is to suggest an approach for estimation of financial risks and forecasting of the client loss and optimal service time utilization based on intellectual data analysis and behavior models.

Methods. To determine the probability of customer loss the neural networks theory, gradient busting, random forest and logistic regression are used. The survival analysis models for possible client transition time to another company are developed.

Results. The best model for forecasting the clients intending for transition to another telecommunication company turned out to be the one based on gradient busting.

Conclusions. It was shown that timely estimation of financial losses, provoked by possible loss of clients, is an urgent task for intellectual data analysis. A perspective approach for optimization of the company financial resources is determining the time period related to possible loss of clients.

Keywords: telecommunication company; financial risks; survival models; gradient busting; neural network; logistic regression.

Вступ

Український ринок телекомунікаційних послуг, який почав стрімко розвиватися наприкінці 90-х років ХХ ст., змінюючи умови для своїх користувачів, пропонуючи дешевші послуги, відмовляючись від плати за вхідні дзвінки тощо, досяг своєї насиченості й охопив більше 90 % українців. Більше того, згідно з дослідженнями, багато українців користуються послугами кількох мобільних операторів, постійно змінюючи їх і шукаючи вигідніші пропозиції, не прив'язуючись до конкретного номера телефону. Щорічно оператори телекомунікаційних послуг несуть значні втрати через відплив абонентів.

Сьогодні у клієнтів телекомунікаційних компаній є можливість вибору з-поміж великої кількості провайдерів послуг. На цьому конкурентному ринку клієнти вибирають високу якість за меншу ціну, в той час як провайдери сконцентровані на створенні вигідних пропозицій. Тому й проблема прогнозування схильності абонента перейти до іншого оператора є дуже актуальною. Крім цього, важливою задачею є не лише виявити абонентів, схильних до зміни мобільного оператора, а й спрогнозувати момент (період) відпливу клієнта. Володіючи

інформацією про таких типових клієнтів, які змінюють операторів періодично (наприклад, кожні 6 місяців тощо), можна запропонувати новий тарифний пакет або якісь додаткові послуги, наприклад безкоштовний інтернет на декілька місяців, і таким чином вплинути на клієнта, щоб він продовжив користуватися послугами певного оператора.

Проведене в 2016 р. компанією TNS Infratest дослідження, що охопило більше 60000 респондентів у 61-й країні світу, оцінювало кількість користувачів мобільних пристроїв та їх уподобання щодо телекомунікаційних послуг. Так, в Україні дослідження проводилося в період з січня по березень 2016 р., вибірка – 1000 респондентів у віці 16+, у форматі телефонного та персонального інтерв'ю. Компанія Google, яка володіє акаунтами користувачів мобільних пристроїв, навела результати дослідження особливостей поведінки українських інтернет-користувачів. Зокрема, порівняно з 2013-м роком кількість користувачів смартфонів в Україні зросла на 150 %. За даними дослідження, вже 35 % українців є користувачами смартфонів, при тому що в 2015 р. ця цифра становила 28 %. Серед молодих людей у віці до 35 років ця цифра ще вища – в цій віковій категорії 72 % українців використовують смартфони [1].

*corresponding author: pbidyuke_00@ukr.net

Згідно з результатами дослідження, в 2016 р. кількість підключених до інтернету пристроїв, що припадає на одного українця, становить у середньому 1,2, що на 58 % більше порівняно з 2013 р.

Ця інформація є цікавою та корисною українським операторам мобільного зв'язку для розробки нових пропозицій для клієнтів, здебільшого орієнтованих на користування інтернетом й альтернативними класичним дзвінком мобільними додатками (Viber, WhatsApp, Telegram) з урахуванням напрямів їх міжнародних дзвінків і використання класичних повідомлень.

Постановка задачі

Метою роботи є моделювання фінансових ризиків методами інтелектуального аналізу даних для прогнозування можливості втрати певного клієнта, а також розробка інноваційного підходу для прогнозування періоду відпливу клієнтів. Прогнозування періоду, коли абонент захоче змінити тарифний пакет або оператора, тобто прогнозування часу відпливу клієнта за допомогою моделей виживання, дає змогу оцінити фінансові ризики телекомунікаційної компанії, пов'язані з недоотриманням очікуваного прибутку через відплив клієнтів. На основі спрогнозованого періоду можливої втрати клієнтів необхідно розробити стратегію поведінки в часі для операторів і додаткових маркетингових компаній з метою забезпечення постійності клієнтів та запобігання плінності клієнтів і можливого недоотримання прибутку.

Опис вхідних даних

Задача прогнозування відпливу абонентів розв'язувалась на фактичних вхідних даних української телекомунікаційної компанії, назва якої не наводиться в роботі з причин зобов'язання авторів не розголошувати конфіденційну інформацію. Вхідна вибірка складалась зі 150 тис. абонентів та, відповідно, інформації про їх активність у мережі протягом 15 місяців (весь 2014 р. і початок 2015 р.). Кожний місяць активності абонента описується такими 10 показниками:

- кількість хвилин вхідних дзвінків (INCOMING);
- кількість хвилин вихідних дзвінків на стаціонарні номери (PSTN);
- кількість хвилин вихідних дзвінків на мобільні номери інших операторів (ALIEN);

- кількість хвилин вихідних дзвінків на мобільні номери цього ж оператора в одному регіоні (REGION);

- кількість хвилин вихідних дзвінків на мобільні номери цього ж оператора в інший регіон (AREA);

- кількість хвилин вихідних дзвінків на мобільні номери інших мобільних операторів (OMO_MINS);

- кількість хвилин вихідних дзвінків на мобільні номери всередині мережі (ONNET_MINS);

- кількість хвилин вихідних дзвінків на міжнародні номери (INTERN_MINS);

- кількість мегабайт спожитого інтернет-трафіку (GPRS_USG_MB);

- кількість надісланих СМС (SMS).

Також відома інформація щодо дати активації абонента (Oblast_Activated), його статі (SEX), віку (AGE), індикатора, чи є він корпоративним клієнтом (COMPANY), а також моделі пристрою зв'язку (мобільний телефон, планшет тощо).

Підготовка даних та їх попередня обробка

Важливим етапом побудови моделі з використанням методів інтелектуального аналізу даних є обробка початкових даних і правильний відбір змінних для навчальної вибірки. Якість побудованої моделі, результат навчання класифікатора суттєво залежать від набору наведених характеристик і сформованих навчальної та перевіркової вибірок.

Попереднє оцінювання характеристик

Спочатку здійснено оцінювання значущості вхідних змінних та виділено найважливіші змінні для аналізу. Для цього використовувались характеристики інформаційного значення (IV) та WOE (Weight of Evidence). Зокрема, з точки зору бізнес-інтересів та оцінених характеристик було виявлено, що сьогодні оператори мобільного зв'язку відходять від так званих "регіональних" тарифів, пропонуючи універсальні тарифні пакети для дзвінків за усіма напрямками домашньої мережі за одним і тим самим тарифом. А тому кількість хвилин вихідних дзвінків на мобільні номери цього ж оператора в цьому регіоні (REGION) чи іншому (AREA) є незначущою змінною. Ще однією характеристикою, яка була виключена на етапі попереднього аналізу, стала кількість хвилин

вхідних дзвінків (INCOMING), оскільки з відміною плати за вхідні дзвінки ця інформація дає змогу лише побачити активність абонента за відсутності користування ним іншими послугами (вихідними дзвінками, інтернетом, повідомленнями тощо).

Крім цього, сформовано зведену характеристику Subscription_Type_Code, яка замінила змінні “Стать клієнта” та “індикатор належності до корпоративного клієнта”. Таким чином, змінна набуває значення m/f для індивідуального клієнта (чоловічої або жіночої статі) та CRP для корпоративного клієнта, що обслуговується за тарифами для бізнесу.

Моделювання за допомогою методу ковзного вікна

Для перетворення задачі передбачення часового ряду на класичну задачу машинного навчання застосовано метод ковзного вікна [2, 3]. Суть методу ковзного вікна така: для кожного вихідного часового ряду $x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, \dots$ вибираються відповідне вікно $w^j = \langle x_{t+1}^{(i)}, x_{t+2}^{(i)}, x_{t+l-1}^{(i)}, x_{t+l}^{(i)} \rangle$ та $y^j = x_{t+l+1}^{(i)}$ – наступний за вікном елемент часового ряду, де t – значення зсуву для вікна, l – фіксована ширина вікна. Зазвичай для кожного початкового часового ряду вибираються кілька послідовних вікон (вікно ковзає часовим рядом) для того, щоб збільшити кількість зразків у підсумковій вибірці. Тоді нова задача має такий вигляд: побудувати модель машинного навчання, яка за $w(j)$ передбачає наступний за вікном елемент $y(j)$. У початковій вибірці даних для кожного абонента відомі 10 показників його активності протягом 15 місяців. Якщо абонент у певному місяці ще не був зареєстрований (активований) оператором, то його активність у цьому місяці дорівнює нулю. Для початкових даних була вибрана ширина вікна в 4 місяці. Для кожного часового ряду $x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, \dots, x_{15}^{(i)}$, де $x_j^{(i)}$ – вектор із 10-ти показників активності за j -й місяць для i -го абонента, зафіксовані одне вікно $w^i = \langle x_{t+1}^{(i)}, x_{t+2}^{(i)}, x_{t+3}^{(i)}, x_{t+4}^{(i)} \rangle$ та сумарна за всіма показниками активність абонента в наступному після вікна місяці $AS(x_{t+5}^{(i)}) = \sum_{k=1}^{10} x_{t+5}^{(i)}[k]$. Абонент вважається таким, що пішов у місяці $t+4$, якщо $AS(x_{t+5}^{(i)}) = 0$, в іншому випадку абонент вважа-

ється утриманим. Якщо для цього абонента можливо вибрати такий час t , що він буде вважатися таким, що пішов, то вибирається саме цей час t . При цьому в підсумкову вибірку попадає тільки частина вікна: всі елементи вікна, крім останнього, тому що для абонентів, які вже пішли, активність у місяці відтоку $t+4$ є набагато меншою, ніж у попередні місяці, і це може стати причиною перенавчання класифікатора.

Додаткові статистичні характеристики. Для покращення моделі машинного навчання стандартним підходом (зокрема, для часових рядів [4]) є підрахунок та включення додаткових характеристик, таких як арифметичне відношення, геометричне відношення, математичне сподівання, дисперсія, коефіцієнти асиметрії та ексцесу.

Додаткові параметри моделі. Для кожного активного місяця абонента у вихідних даних надано код пристрою (Type Allocation Code (TAC)), тобто перші 8 цифр IMEI, що описують модель і місце походження пристрою. Для кожного абонента з вибірки зафіксовано один TAC-код, що зустрічався найбільшу кількість разів у досліджуваному вікні; якщо таких кодів кілька, тоді вибирався останній у хронологічному порядку. За цим кодом TAC отримані такі додаткові характеристики про пристрої зв'язку:

- працює на Android (бінарна характеристика);
- працює на iOS (бінарна характеристика);
- працює на Windows Phone (бінарна характеристика);
- підтримує передачу даних GPRS (бінарна характеристика);
- підтримує передачу даних EDGE (бінарна характеристика);
- підтримує передачу даних 3G (бінарна характеристика);
- має модуль Wi-Fi (бінарна характеристика);
- рік випуску моделі;
- коефіцієнт вартості від 1 до 10;
- діагональ екрана.

Вибрані класифікатори для розв'язання прикладної задачі. Слід зазначити, що пошук публікацій щодо розв'язання такого типу задачі, як прогнозування відтоку мобільних абонентів, не дав жодних результатів. Така задача є новою, а тому автори вирішили відібрати самостійно моделі й методи машинного навчання, які здавалися більш перспективними для цих типів даних, і виконати прогнозування за їх допомогою. Відібрано такі моделі: логістичну регресію, ней-

ронні мережі, випадковий ліс, градієнтний бустинг над деревами рішень.

Підбір параметрів. Для всіх моделей машинного навчання перед початком самого навчання необхідно визначити параметри моделі. Для логістичної регресії це параметр регуляризації; для нейронних мереж – це кількість нейронів у скритому шарі; в методі випадкового лісу – це кількість дерев; у градієнтному бустингу – це кількість дерев тощо. Виникає запитання, як вибрати параметри моделі машинного навчання для отримання найкращих результатів? Для цього використано алгоритм пошуку на сітці (Grid search) [5–7] (рис. 1).

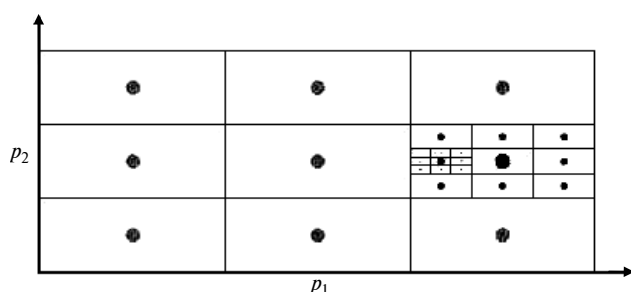


Рис. 1. Ілюстрація роботи алгоритму пошуку на сітці при $n = 2$

Для кожного з n параметрів моделі визначено інтервали пошуку $[a_i, b_i]$, $1 \leq i \leq n$, у яких знаходяться всі допустимі значення. Уявімо фіксовані n значень як точку в n -вимірному кубі. Суть алгоритму полягає в тому, щоб розбити цей куб на куби меншого розміру, в кожному з яких вибирається центральна точка p_j . Далі виконується навчання моделі з параметрами p_j для кожного з j кубів. Потім обчислюється метрика якості навчання у кожній з j точок та вибирається найкраща точка. З тим кубом, що відповідає найкращій точці, необхідно зробити те ж саме, і так далі до тих пір, поки не знайдемо ту точку, що нас задовольняє. Цей алгоритм не завжди знаходить оптимальні параметри моделі машинного навчання, в деяких випадках він знаходить тільки локально оптимальні параметри. Щоб уникнути цього, використовується дрібніше розбиття куба на елементи.

Таблиця. Порівняння моделей на тестовій вибірці

Модель	Precision	Recall	F1	F0,5	AUC
Градієнтний бустинг, GB	0,70	0,64	0,66	0,68	0,842 ± 0,005
Випадковий ліс, RF	0,72	0,60	0,65	0,69	0,832 ± 0,008
Нейронні мережі, NN	0,69	0,59	0,63	0,66	0,825 ± 0,007
Логістична регресія, LR	0,63	0,37	0,46	0,55	0,842 ± 0,002

Підготовка вибірки

Після етапу обробки даних отримано вибірку з 96460 зразків із 84 характеристиками; вона була поділена на такі три частини:

- навчальна вибірка (60 %);
- тестова вибірка (20 %);
- крос-валідаційна вибірка (20 %).

На тестовій вибірці виконується передбачення за моделлю та оцінюються результати за заданими метриками. Найкраща модель вибирається за результатами передбачень за тестовою вибіркою. Крос-валідаційна вибірка використовується для незалежної оцінки моделей.

Побудова моделей

Вибірка даних була випадковим чином перемішана на три частини (60, 20, 20 %) 10 разів, після кожного разу виконувалося навчання моделей і перевірка моделі на тестовій вибірці. Зокрема, для моделей логістичної регресії та нейронних мереж характеристики отриманої вибірки були масштабовані на відрізок $[0; 1]$. Результати використання моделей наведені в таблиці.

Інтерпретація результатів. Підсумкові порівняння класифікаторів зроблені за допомогою параметра якості AUC (площа під ROC-кривими, побудованими за результатами аналізу і проілюстрованими на рис. 2). Найкращі результати показав градієнтний бустинг над деревами рішень, схожі з цим результати показав класифікатор “випадковий ліс”. Однак у задачі відтоку абонентів також важливою метрикою є загальна точність (precision). Передбачення відтоку абонента робиться з метою запропонувати клієнту вигідніші умови, бонуси і таким чином втримати його. Саме точність показує, наскільки точно ці пропозиції будуть надходити абоненту, якому це необхідно. За низької точності класифікатора мобільний оператор часто відправлятиме пропозиції клієнтам, які не планують переходити від одного оператора до іншого, що може призвести до фінансових втрат.

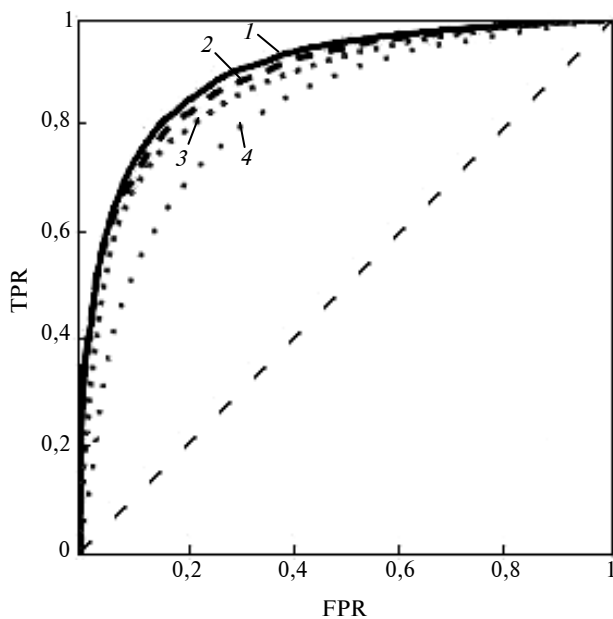


Рис. 2. ROC-криві побудованих моделей: 1 – GB; 2 – RF; 3 – NN; 4 – LR

Наступним важливим моментом для телекомунікаційної компанії є можливість напрацювання стратегії вчасного запобігання відтоку абонентів. Для цього можна розробити спеціальні акційні пропозиції і направити їх за декілька днів або тижнів до прийняття абонентом рішення щодо зміни оператора. Для цього необхідно спрогнозувати час можливого відтоку клієнтів і визначити категорію клієнтів, які мають схильність до зміни оператора, щоб оптимізувати витрати телекомунікаційної компанії, в т.ч. на розсилку акційних пропозицій лише цільовим клієнтам. Для цього використаємо підхід теорії аналізу виживання для оцінювання часу можливого відтоку клієнтів.

Теоретичні положення теорії виживання для аналізу ризиків

Теорія виживання отримала відповідну назву внаслідок її широкого застосування в медичних дослідженнях для оцінки тривалості життя при вивченні ефективності методів лікування. Пізніше ці методи стали застосовуватись у страховій сфері, а також у соціальних науках. Аналіз виживання [7–9] займається моделюванням процесів настання термінальних (критичних) подій для елементів

тієї чи іншої сукупності (спочатку – “смерті” для елементів сукупності живих істот). Так, у рамках медичних досліджень аналіз виживання може відповідати на такі запитання: “Яка буде частка тих, що вижили, серед пацієнтів через деякий час після застосованих методів лікування?”, “Які темпи смертності будуть спостерігатися серед тих, що вижили?”, “Які чинники впливають на збільшення або зменшення шансів на виживання?” тощо. Щоб відповісти на ці запитання, необхідно мати можливість чітко визначити “час життя” елемента (період перебування елемента в сукупності з настанням термінальної події). У випадку з біологічним виживанням “смерть” однозначна, але в інших випадках настання термінальної події не завжди можливо локалізувати в окремий момент часу.

У цілому аналіз виживання орієнтований на побудову моделей, що описують дані про час настання події. Оскільки живий організм може померти лише один раз, то традиційно в рамках цього підходу розглядаються лише поодинокі й одноразові термінальні події.

Цензурування змінних

Аналіз даних методами аналізу виживання може бути здійснений тільки для цензурованих даних. Спостереження називаються цензурованими, якщо спостережувана залежна змінна відображає момент настання термінальної події, а тривалість дослідження обмежена в часі. Можливі механізми цензурування змінних: фіксоване цензурування (спостереження відбувається протягом фіксованого проміжку часу) та випадкове цензурування (спостереження відбувається протягом проміжку часу, який настає після того моменту, коли елементи вибірки пережили певну подію).

При цензуруванні можна вказати напрям, у якому проводиться цензурування. Сьогодні існують такі напрями цензурування: правостороннє цензурування (відомо, в який момент експеримент був розпочатий і що він закінчиться в момент часу, розташований праворуч від точки початку експерименту), лівостороннє цензурування (якщо не має інформації про те, коли експеримент було розпочато) та інтервальне цензурування (коли відома лише інформація, що час виживання розподілений між значеннями a і b ($t \in [a, b]$) [10, 11].

Таблиці виживання

Традиційним методом дослідження виживання є таблиці виживання. Така таблиця є таблицею частот, де область дослідження розбивається на деяку кількість інтервалів, обумовлених особливостями дослідження. Для кожного інтервалу обчислюється кількість об'єктів, які на початку розглянутого інтервалу не мали певної ознаки, і кількість об'єктів, які набули певної ознаки в цьому інтервалі. Потім розраховуються відносні частки цих об'єктів, на основі яких оцінюється ймовірність події (щільність імовірності, кумулятивна вірогідність), функція ризику події (hazard ratio), а також їх стандартні похибки.

Також обчислюється кількість об'єктів, вилучених або цензурованих на кожному інтервалі [8].

Функція виживання

Аналізований об'єкт у функції виживання традиційно умовно позначається S і описується такою функцією:

$$S(t) = P(T > t),$$

де t – це деякий час, у ході якого виконувалося спостереження за сукупністю; T – випадкова величина, що позначає момент "смерті" (покидання об'єктом сукупності); P означає ймовірність "смерті" у заданому часовому інтервалі [9]. Тобто функція виживання описує вірогідність "смерті" через якийсь час після моменту t .

Зазвичай передбачається, що на початковий момент $S(0) = 1$, хоча це значення може бути і меншим, якщо задачею передбачена можливість негайної смерті або невдачі.

Прогнозування періоду відтоку клієнтів за теорією виживання

Для прогнозування можливого часу відтоку зроблена вибірка зі 141430 спостережень, далі були відібрані (відцензуровані) ті клієнти, які переходили в іншу компанію в цей період. У результаті було відцензуровано 86915 клієнтів та побудовано таблицю виживання. За результатами побудови таблиць виживання та моделі виживання (рис. 3) можна побачити, що протягом перших двох місяців (60 днів) відток клієнтів зовсім не спостерігається і ймовірність виживання становить 1. Перший відтік спосте-

рігається приблизно через 3 місяці, що можна пояснити закінченням акційних пропозицій після активації пакета. Далі з часом ймовірність відтоку зростає і через півроку становить 0,2, через рік – 0,3 і т. д.

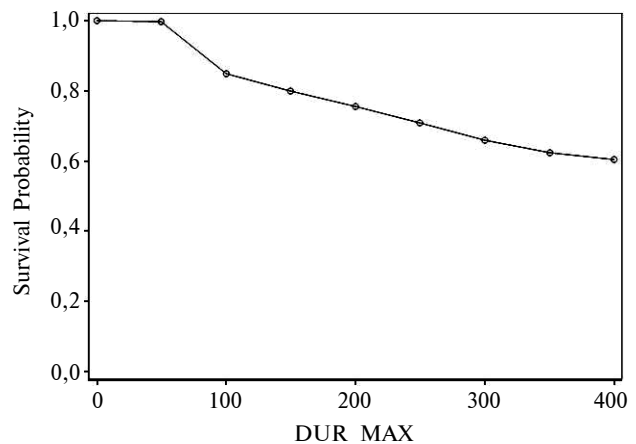


Рис. 3. Функція ризику виживання, побудована за таблицею виживання

Наступний етап задачі аналізу фінансових ризиків – оцінювання можливих втрат телекомунікаційної компанії через вплив певної частки клієнтів. Для цього побудована й оцінена функція рівня ризику залежно від часу обслуговування клієнта (рис. 4). Тут також підтверджується, що найбільший рівень ризику спостерігається через 3 місяці. Сама частка клієнтів, що здійснює відтік, є невеликою у загальній вибірці, тому й hazard rate становить 0,0032.

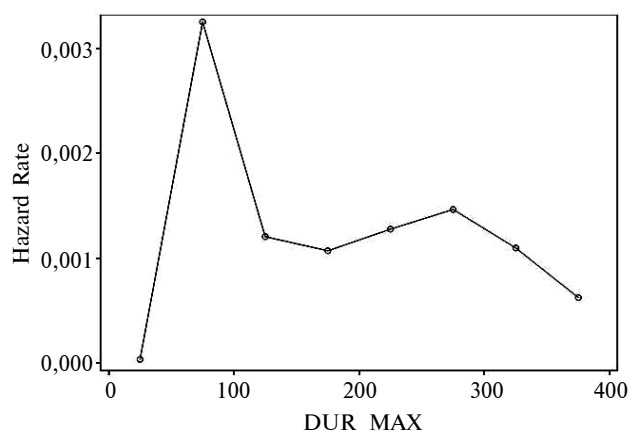


Рис. 4. Функція ризику в часі

Висновки

Виконане дослідження – одне з перших, спрямованих на аналіз можливості оцінювання фінансових ризиків методами інтелектуального

аналізу даних у телекомунікаційній сфері. Отримані результати свідчать про високу предикативну здатність моделей, побудованих для передбачення ймовірності відтоку клієнтів. Такий підхід може бути взятий за основу для CRM-галузі (управління взаємовідносинами з клієнтами) як найбільш чутливої та залежної від уподобань клієнтів. Крім того, вперше запропоновано використання моделей, побудованих на основі теорії виживання для оцінювання можливого критичного моменту часу, коли починається вплив клієнтів.

Запропонований підхід планується розширити на прогнозування можливих втрат компанії з урахуванням часу та побудови моделей виживання складнішого типу, наприклад пропорційних ризиків, параметричних та непараметричних нелінійних моделей. Також доцільно розв'язати задачу комбінування результатів оцінювання, отриманих за допомогою різних методів. Як правило, такий підхід дає можливість підвищити якість остаточного результату завдяки додатковій компенсації похибок класифікації.

Список літератури

1. Google з'ясував, як українці користуються інтернетом [Електронний ресурс]. – <http://watcher.com.ua/2016/09/13/google-z-yasuvav-yak-ukrayintsi-korystuyutsya-internetom/>
2. Воронцов К. Математические методы обучения по прецедентам. Курс лекций. – М.: МФТИ, 2006. – 58 с.
3. Бідюк П.І., Романенко В.Д., Тимошук О.Л. Аналіз часових рядів. – К.: НТУУ “КПІ”, 2013. – 600 с.
4. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: Изд-во Ин-та математики, 1999. – 61 с.
5. Чубукова И.А. Data Mining. – М.: Бинум ЛБЗ, 2008. – 384 с.
6. Довгий С.О., Бідюк П.І., Трофимчук О.М. Системи підтримки прийняття рішень на основі ймовірнісно-статистичних методів. – К.: Логос, 2014. – 419 с.
7. Загоруйко Н.Г., Ёлкина В.Н., Лбов Г.С. Алгоритмы обнаружения эмпирических закономерностей. – Новосибирск: Наука, 1985. – 120 с.
8. Cox D.R. Regression models and life-tables // J. Royal Statist. Soc. Ser. B (Methodological). – 2007. – 34, № 2. – P. 187–220.
9. Фомін О.В., Кузнєцова Н.В. Скорингові моделі поведінки клієнтів-власників кредитних карток для оцінки їх платоспроможності // Системні науки та кібернетика. – 2016. – № 5. – С. 56–67.
10. Marimo M. Survival analysis of bank loans and credit risk prognosis [Online]. – Available: http://wiredspace.wits.ac.za/jspui/bitstream/10539/18597/1/Mercy%20Marimo%20Thesis_Survival%20Analysis_28.03.%202015_v1.pdf
11. Кузнєцова Н.В., Куца К.В., Штогрін С.Р. Застосування методології аналізу виживання для дослідження споживчих ризиків // Системні науки та кібернетика. – 2017. – № 6. – С. 126–135.

References

- [1] Google Determined the Way Ukrainians use Internet [Online]. Available: <http://watcher.com.ua/2016/09/13/google-z-yasuvav-yak-ukrayintsi-korystuyutsya-internetom/> (in Ukrainian).
- [2] K. Vorontsov, *Mathematical Learning Methods on Precedents. Course of Lectures*. Moscow, Russia: MPTI, 2006 (in Russian).
- [3] P.I. Bidyuk et al., *Analysis of Time Series*. Kyiv, Ukraine: NTUU KPI, 2013 (in Ukrainian).
- [4] N.G. Zagoruyko, *Applied Data and Knowledge Analysis Methods*. Novosibirsk, Russia: Institute of Mathematics Publ., 1999 (in Russian).
- [5] I.A. Chubukova, *Data Mining*. Moscow, Russia: Binom LBZ, 2008.
- [6] S.O. Dovgij et al., *Decision Making Systems on Probability-Statistic Methods*. Kyiv, Ukraine: Logos, 2014 (in Ukrainian).
- [7] N.G. Zagoruyko et al., *Algorithms of Empirical Patterns Detection*. Novosibirsk, SU: Nauka, 1985 (in Russian).
- [8] D.R. Cox, “Regression models and life-tables”, *J. Royal Statist. Soc. Ser. B (Methodological)*, vol. 34, no. 2, pp. 187–220, 2007.
- [9] O.V. Fomin and N.V. Kuznietsova, “Scoring models of credit card holder behavior for their solvency estimation”, *Systemni Nauky ta Kibernetika*, no. 5, pp. 56–67, 2016 (in Ukrainian).
- [10] M. Marimo. (2005, Apr. 12). *Survival Analysis of Bank Loans and Credit Risk Prognosis* [Online]. Available: http://wiredspace.wits.ac.za/jspui/bitstream/10539/18597/1/Mercy%20Marimo%20Thesis_Survival%20Analysis_28.03.%202015_v1.pdf
- [11] N.V. Kuznietsova et al., “Using the Methodology of Survivability analysis for consumer risks researching”, *Systemni Nauky ta Kibernetika*, no. 6, pp. 126–135, 2017 (in Ukrainian).

Н.В. Кузнєцова, П.І. Бідюк

МОДЕЛЮВАННЯ ФІНАНСОВОГО РИЗИКУ В ТЕЛЕКОМУНІКАЦІЙНІЙ СФЕРІ

Проблематика. Телекомунікаційна сфера в Україні динамічно розвивається, постійно оновлюючи пропозиції під вимоги ринку і споживача. Тому сьогодні особливо актуальною задачею є своєчасне оцінювання фінансових ризиків, оптимізація фінансових витрат, пов'язаних із появою нових компонентів і можливим відтоком клієнтів до конкурентів.

Мета дослідження. Запропонувати підхід до оцінювання фінансових ризиків, прогнозування відтоку клієнтів та оптимального часу користування послугами на основі моделей інтелектуального аналізу даних і поведінкових моделей.

Методика реалізації. Застосовано нейронні мережі, градієнтний бустинг, випадковий ліс і логістичну регресію для визначення ймовірності відтоку клієнтів. Розроблено моделі аналізу виживання для визначення можливого часу переходу клієнта до іншої компанії.

Результати дослідження. Для прогнозування клієнтів, які збираються змінити телекомунікаційну компанію, найкращою виявилась модель на основі градієнтного бустингу.

Висновки. Вчасне оцінювання фінансових витрат, пов'язаних із можливим відтоком клієнтів, є актуальною задачею, що може розв'язуватися з використанням різних методів інтелектуального аналізу даних. Перспективним підходом є визначення періоду, в який можливим є відтік клієнтів, для оптимізації фінансових ресурсів компанії.

Ключові слова: телекомунікаційна компанія; фінансові ризики; моделі виживання; градієнтний бустинг; нейронна мережа; логістична регресія.

Н.В. Кузнєцова, П.І. Бідюк

МОДЕЛИРОВАНИЕ ФИНАНСОВОГО РИСКА В ТЕЛЕКОММУНИКАЦИОННОЙ СФЕРЕ

Проблематика. Телекоммуникационная сфера в Украине динамически развивается, постоянно обновляя предложения под требования рынка и потребителя. Поэтому особенно актуальной задачей сегодня являются своевременное оценивание финансовых рисков и оптимизация финансовых расходов, связанных с появлением новых компонент и возможным оттоком клиентов к конкурентам.

Цель исследования. Предложить подход к оцениванию финансовых рисков, прогнозированию оттока клиентов и оптимального времени пользования услугами на основе моделей интеллектуального анализа данных и поведенческих моделей.

Методика реализации. Используются нейронные сети, градиентный бустинг, случайный лес и логистическая регрессия для определения вероятности оттока клиентов. Разработаны модели анализа выживания для определения возможного времени перехода клиента в другую компанию.

Результаты исследования. Для прогнозирования клиентов, которые собираются поменять телекоммуникационную компанию, лучшей оказалась модель на основе градиентного бустинга.

Выводы. Своевременное оценивание финансовых потерь, связанных с возможным оттоком клиентов, является актуальной задачей интеллектуального анализа данных. Перспективным подходом является определение периода времени, в который возможен отток клиентов, для оптимизации финансовых ресурсов компании.

Ключевые слова: телекоммуникационная компания; финансовые риски; модели выживания; градиентный бустинг; нейронная сеть; логистическая регрессия.

Рекомендована Радою
Навчально-наукового комплексу
"Інститут прикладного системного
аналізу"
КПІ ім. Ігоря Сікорського

Надійшла до редакції
30 червня 2017 року