В.М. Бредіхін, Т.С. Сенчук, К.С. Стужук

Харківський національний університет міського господарства імені О.М. Бекетова, Україна

ФОКУС НА ШТУЧНИЙ ІНТЕЛЕКТ ДЛЯ ПРОГНОЗУВАННЯ ВІДТОКУ КЛІЄНТІВ З САЙТІВ ON-LINE ОСВІТИ

У статті проведено дослідження процесу прогнозування відтоку клієнтів, що має особливо велике значення для компаній, які використовують бізнес-модель на основі підписки. Було встановлено, що показник відтоку вкрай важливий для компаній з передплатною та транзакційною моделлю бізнесу, які мають на увазі регулярні платежі у бік компанії (банки, оператори зв'язку, SaaS-сервіси тощо). Для цього було розглянуто типи, основні причини відтоку клієнтів та параметри визначені для побудови прогнозової моделі з використанням алгоритмів машинного навчання. Результатом цього стало побудування гіпотез причин відтоку клієнтів з сайтів що надають послуги навчання на основі курсів, які представлено on-line в internet просторі. Для побудови моделі прогнозування відтоку було вивчено як поведінкові особливості студентів, їх мотивацію та структуру самих курсів. На основі зібраного великого масиву даних, було проаналізовано за великою кількістю параметрів їх зміна та виявлено залежності між поведінковими особливостями студентів, структурами курсів та їх проходженням. Було побудовано варіант прогнозуючої моделі, для якої потім збільшили точність її роботи та інтегрували результати в модуль передбачення відтоку клієнтів. У підсумковий список ознак увійшли понад 100 параметрів, які розділили на 6 блоків. За результатом цього створено прогнозну модель з використанням розподілу Вейбула, оскільки поведінку клієнта можна розглядати як свого роду модель виживання. Для оцінки ймовірності відтоку клієнтів, на підставі розглянутих гіпотез була розроблена рекурентна нейронна мережа з LSTM-шаром, де як функцію втрат використано негативну логарифмічну функцію правдоподібності для розподілу Вейбулла. Як висновок було запропоновано впровадження стабільного проактивного освітнього бізнесу, коли рішення приймаються не тільки на основі відчуттів, а й на основі даних, приходить більш чітке та обґрунтоване розуміння того, як покращувати освітній продукт.

Ключові слова: відток клієнтів, транзакційна модель бізнесу, розподіл Вейбула, нейронна мережа, алгоритми машинного навчання.

Постановка проблеми

Відтік клієнтів – це не просто показник у звіті, це справжня проблема, яка призводить до втрати грошей та труднощів у розвитку бізнесу.

Для деяких сфер діяльності поняття відтоку не застосовується (наприклад, продаж нерухомості), оскільки покупки не мають регулярний характер. Показник відтоку вкрай важливий для компаній з передплатною та транзакційною моделлю бізнесу, які мають на увазі регулярні платежі у бік компанії (банки, оператори з'язку, SaaS-сервіси тощо) [1].

У цій статті, щоб вирішити наведені вище проблеми, ми пропонуємо розглянути вирішення цієї проблеми з боку сервісів, що надають освітні послуги on-line.

Аналіз останніх досліджень і публікацій

Компаній із стран СНГ та іноземні компанії (наприклад, SAP, Microsoft, IBM та ін.) надають як готові інструменти для клієнтської аналітики, так і розробляють індивідуальні рішення під окремих

замовників. Найбільші банки, телеком-оператори, ритейлери мають у своєму арсеналі інструменти аналізу великих даних і розробляють власні рішення щодо аналізу та прогнозу відтоку клієнтів (наприклад, МТС, Ощадбанк, ПриватБанк).

Як правило, існуючі розробки ґрунтуються на особистісних даних клієнта, а також на даних про його активність у компанії: послуги та продукти, якими він користується, історія транзакційної активності, історія звернень, інформація про покупки тощо. Отримані дані є великими масивами зі структурованою та неструктурованою інформацією, для аналізу та виявлення прихованих закономірностей, яких широко використовується інтелектуальний аналіз даних та засновані на ньому методи машинного навчання [2].

Мета статті

Мета роботи полягає в тому, щоб на основі набору даних з сайту дистанційних курсів і сучасних методів машинного навчання побудувати прогнозуючу модель відтоку споживачів та завчасно

запропонувати заходи для зменшення цього показника.

Виклад основного матеріалу дослідження

Щодня ринок поповнюється все новими конкурентами, які готові зробити будь-які зусилля, щоб відвоювати частину клієнтури в інших.

Прогнозування відтоку клієнтів має особливо велике значення для компаній, які використовують бізнес-модель на основі підписки. До такого типу організацій належать мобільні оператори, оператори кабельного телебачення та компанії, що обслуговують прийом платежів за допомогою кредитних карток до освітніх сайтів.

Існує три типи відтоку:

- природний – зміна місця проживання, смерть, зміна потреб;
- мотивований – клієнт відмовився від послуг компанії на користь іншої компанії або знайшов альтернативний варіант задоволення потреби;
- прихований – клієнт продовжує користуватися послугами компанії, але обсяг та регулярність споживання знижується. Це вірна ознака того, що клієнт паралельно користується послугами конкурентів чи продуктами-замінниками [3].

Причини відтоку клієнтів різняться залежно від промисловості та специфіки компанії. Тим не менш, низка причин відтоку покупців є універсальними для всіх сфер діяльності:

- невдоволення якістю товарів та послуг компанії – як показують численні дослідження, це головна причина відтоку;
- низький рівень клієнтського сервісу – важливо як запропонувати покупцю якісний продукт, а й забезпечити належний рівень обслуговування;
- найбільш привабливі пропозиції конкурентів, але не обов'язково вигідніші. Сприйнята цінність може серйозно відрізнятись від реальної цінності продукту чи послуги;
- поява продуктів чи послуг замінників – поява на ринку нових рішень, що дозволяють задовольнити потреби клієнтів;
- зниження лояльності до бренду – бажання спробувати щось нове, віяння моди, рекомендації рідних та друзів є досить поширеними причинами відтоку;
- персональні чинники – зміна рівня доходу, сімейного стану, уподобань, хобі тощо [3].

Часто маркетологи для прогнозування відтоку клієнтів використовують формулу:

$$Chart = \frac{Customer\ Lost\ During\ Period}{Customers\ At\ Beginning\ Of\ Period} \quad (1)$$

де: *Customer Lost During Period* – кількість покупців, які залишили вас за період;

Customers At Beginning Of Period – загальна кількість покупців на початок періоду [4].

Це описативна модель відтоку. Вона дає зрозуміти, що сталося, але її можливості обмежені у разі, якщо нам цікава тенденція зміни показника.

Для більш ретельного дослідження відтоку та особливостей поведінки клієнтів, залучених у різний час, використовується когортний аналіз. Він має на увазі поділ бази на когорти – наприклад, залежно від дати першої покупки чи передплати. Далі ви ділите період, що вас цікавить, на інтервали (тижні, місяці, квартали) і простежуєте динаміку дій кожної групи.

Когортний аналіз відтоку дозволяє виявити взаємозв'язок між ставленням клієнтів до вашого продукту та впроваджуваними рішеннями.

Дослідження в галузі побудови прогнозуючих моделей ведуться фахівцями різних галузей, і вже існує безліч підходів і методів до реалізації моделей аналізу даних.

Алгоритми машинного навчання можуть допомогти виявити незайняті сегменти клієнтів, які збираються піти чи піти до конкурента.

Інструменти цієї категорії на базі штучного інтелекту можуть допомогти у збиранні даних, побудові прогнозної моделі, а також тестуванні та перевірці цієї моделі на реальних клієнтах. Ця інформація може вказувати на те, на якій стадії відтоку знаходиться людина. У той час як швидко відштовхуючих клієнтів (користувачів, які відмовляються від продукту незабаром після того, як вони почали його використовувати) важко повторно залучити, клієнтів з пізнім відтоком (тих, хто має тривалі відносини з вашим брендом) можна стимулювати продовжувати використовувати ваш продукт.

Алгоритми штучного інтелекту створюють портрети клієнтів на основі мільярдів точок даних, у тому числі:

- географічні події;
- взаємодія на сайті;
- напрямок джерела;
- психографічні чинники;
- купівельна поведінка;
- минуле спілкування [5].

В результаті алгоритми машинного навчання можуть:

- більш чітко визначити, які клієнтські сегменти слід включати та виключати з кампаній;
- краще підбиравати клієнтам продукти, які вони, швидше за все, використовуватимуть і зберігатимуть;
- запобігти просуванню обмеженого асортименту серед покупців, які неминуче повертають товари.

Основні труднощі полягають в тому, що в моменті неможливо передбачити, хто зі студентів пройде курс до кінця, хто піде. Факт втрати помітний лише тоді, коли студенти перестали заходити на платформу, дивитися уроки та виконувати завдання. А зараз уже не спрацюють ніякі розсилки, повідомлення, акції та персональні пропозиції».

Щоб мінімізувати відтік із курсів, важливо:

- зрозуміти, що може стати причиною відтоку;
- автоматизувати аналітику навчання;
- використовувати інструменти прогнозування.

Що може спричинити відплив студентів з курсу?

Для догляду є чітка причина. Виявити цю причину можна у різний спосіб:

- провести опитування студентів, щоб зрозуміти, що їм подобається в курсі, а що ні;
- оцінити якість зворотного зв'язку про програму;
- проаналізувати дані щодо процесу навчання (швидкість проходження уроків, кількість питань, що задаються, результати проходження тестів і виконання завдань).

Які дані для цього необхідно зібрати:

- збір загальних даних;
- збір даних про перегляд контенту;
- оцінки уроків учасниками курсу;
- новини, що автоматично формуються;
- вільно формовані новини.

Для побудови моделі прогнозування відтоку необхідно вивчити як поведінкові особливості студентів та його мотивацію, а й самі курси [6].

Зібравши великий масив даних, можна проаналізувати їх за великою кількістю параметрів та виявити цікаві залежності між поведінковими особливостями студентів, структурами курсів та їх проходженням.

Докладно ми розглянули цільову особливість, фактичний відтік. За результатами бачимо, що він становив 26,5% від загального обсягу відтоку клієнтів, рис. 1.

Рис. 1. Процент відтоку слухачів з курсів

Також слід звернути увагу, що найбільше слухачів залишають навчання під час проходження першої третини програми, рис. 2.

Рис. 2. Відток слухачів під час проходження першої третини програми курсів

Наступною особливістю є те, що більший відтік спостерігається на курсах, що мають конкретну дату закінчення. Якщо курс немає закінчення, то студент може до нього «повертатися» необмежену кількість разів, рис. 3.

Рис. 3. Відток слухачів з курсів які мають конкретну дату закінчення навчання

Гіпотеза про відтік серед студентів із низькими балами також знайшла своє підтвердження, рис. 4.

Рис. 4. Відток студентів з низькими балами

На основі даних ми побудували перший варіант прогнозуючої моделі. Потім збільшили точність її роботи та інтегрували результати в модуль передбачення відтоку клієнтів.

У підсумковий список ознак увійшли понад 100 параметрів, які розділили на 6 блоків:

- ознаки студента – вся інформація про нього;
- ознаки курсу – загальні характеристики курсу: тривалість, вартість, сертифікат, прохідний

бал та інші;

- ознаки уроків – структура курсу, кількість уроків та його типи;

- ознаки тестових завдань – середня довжина питань, ставлення необхідні проходження тесту балів до кількості балів, метод оцінки результатів тесту (за кількістю правильних відповідей чи з набраним балам);

- ознаки питань тестів – середня довжина відповідей, наявність питань із відкритою відповіддю, з множинним вибором та інші;

- ознаки модерації – швидкість та якість зворотного зв'язку з питань з відкритими відповідями.

Одна з найцікавіших особливостей, які стали основою рекомендацій для складання курсів, – це реакція студентів на різні питання до тестів.

При складанні курсу слід враховувати, що чим більше тестів із вільними відповідями, які потребують перевірки куратором, тим вищий відтік. А також на відтік негативно впливає велика кількість питань із множинними відповідями, рис. 5.

Рис. 5. Реакція студентів на різні питання до тестів

Тестування моделі дало високі результати точності і дозволило вирішити кілька основних завдань:

- чітке розуміння, які ознаки характерні для студентів з високою ймовірністю відтоку;

- дозволило виявляти студентів, схильних до відтоку та застосовувати превентивні заходи для їх утримання.

Як сімейство розподілів, що моделюють час відтоку, було використано розподіл Вейбула, оскільки поведінку клієнта можна розглядати як свого роду модель виживання.

Залежно від параметрів розподілу густини ймовірності Вейбула графіки представлено на рис. 6.

Це щільність розподілу ймовірності відтоку клієнта трьох різних груп із часом. Час представлений у місяцях. Іншими словами, цей графік показує, коли найімовірніше відбудеться відтік

клієнта в наступні два місяці. Як видно, клієнт з розподілом має великий потенціал піти раніше, ніж клієнти з розподілами Вейбула (2, 0.5) та Вейбула (3,1).

Рис. 6. Щільність розподілу ймовірності відтоку клієнта трьох різних груп із часом

Для оцінки ймовірності відтоку клієнтів, на підставі розглянутих гіпотез була розроблена рекурентна нейронна мережа з LSTM-шаром, де як функцію втрат використовувалася негативна логарифмічна функція правдоподібності для розподілу Вейбулла.

Цей підхід розглядався на тимчасовому лазі 2 місяці, але його можна масштабувати при розбитті часових проміжків на дрібніші (наприклад, при розбитті місяців на тижні).

В основі метрики оцінки якості моделі вибрали Lift Curve. Вона визначає у скільки разів алгоритм робить краще прогноз на перших 30% даних, ніж випадковим чином».

Тестова вибірка складалася з 500 заздалегідь відібраних клієнтів, яких не було в навчальній вибірці. Для моделі проводився підбір гіперпараметрів з використанням крос-валідації з розбивкою по клієнтам. Для навчання кожної моделі використовувалися однакові набори ознак.

Висновки

Як висновок слід відзначити, що відтік клієнтів найпростіше можна простежити у галузях, що надають свої послуги на основі регулярної оплати за певний період, де відтік бази означає кількість користувачів, які не продовжили підписки.

Для ресурсів, що пропонують користувачам контент, виражається у кількості відписок. Відтік клієнтів для магазину – це кількість покупців, які не здійснили повторні покупки. У торгівлі досліджувати процес відтоку стає трохи складніше, тому що не зовсім зрозуміло, чи можна вважати клієнта втраченим, чи він все ж таки повернеться за покупкою через деякий час. Зазвичай в такому випадку використовують аналіз за частотою покупок: якщо користувач не здійснив покупку в певний момент, найімовірніше, він не зробить її зовсім.

Технологія машинного навчання дає величезні перспективи створення стабільного проактивного освітнього бізнесу, коли рішення приймаються не тільки на основі відчуттів, а й на основі даних, приходить більш чітко та обґрунтоване розуміння того, як покращувати освітній продукт. Про це також кажуть міжнародні прогнози.

А ще це великий внесок у лояльність та ефективність викладачів: тепер їм не потрібно самостійно здогадуватися про причини поганих результатів, вони можуть заздалегідь отримувати сигнали та вирішувати проблеми до їхньої появи.

Література

1. Клименко Н.А. Оцінка ризиків на ринку послуг мобільних операторів / Клименко Н.А., Вороненко І.В., Нагорна О.В., Громик Н.В. // Ефективна економіка. – 2021. – №7. DOI: <https://doi.org/10.32702/2307-2105-2021.7.92>
2. Кос І. Сучасні інструменти управління консалтинговою компанією // Інноваційна Україна: креативні ідеї та проекти : зб. доп. 87-ї наукової студентської конференції, 4–13 травня 2020 р. – Київ, КНЕУ, 2020. – С. 76–78.
3. Відтік клієнтів: причини та усунення [Електронний ресурс]. – Режим доступу: <https://fractus.com.ua/uk/blog/korysni-statti/prodazhi/vidtik-kliientiv-prichini-ta-usunennja/>
4. Основні терміни та метрики – інтернет маркетингу, веб аналітики та контекстної реклами [Електронний ресурс]. – Режим доступу: <https://bizautomation.com.ua/internet-marketing-terminu/>
5. Загарчук В.І. Використання штучного інтелекту у торгівлі / Загарчук В.І., Клим О.С., Антохова І.М. // International scientific e-journal АГОС. ONLINE. – 2020. – № 15. DOI: <https://doi.org/10.36074/2663-4139.15.10>
6. Стужук К.С. Штучний інтелект для прогнозування відтоку клієнтів / Стужук К.С., Бредіхін В.М. // Перспективи розвитку територій: теорія і практика : матер. VI Всеукр. наук.-практ. конф. здобувачів вищої освіти і молодих учених, 18–19 листопада 2021 р.

References

1. Klymenko, N., Voronenko, I., Nahorna, O., Gromyk, N. (2021). Risk assessment in the market of services of mobile operators. *Efektivna ekonomika*, 7. DOI: <https://doi.org/10.32702/2307-2105-2021.7.92> [in Ukrainian]
2. Kos, I. (2020). Modern management tools for a consulting company. *Proceedings of the 87th Scientific Student Conference “Innovative Ukraine: creative ideas and projects”*,

4–13 May 2020. Kyiv, KNEU, 76–78 [in Ukrainian]

3. Customer outflow: causes and remedies. URL: <https://fractus.com.ua/uk/blog/korysni-statti/prodazhi/vidtik-kliientiv-prichini-ta-usunennja/> [in Ukrainian]

4. Basic terms and metrics - internet marketing, web analytics and contextual advertising. URL: <https://bizautomation.com.ua/internet-marketing-terminu/> [in Ukrainian]

5. Zagarchuk, V., Klym, O., Antokhova, I. (2020). The use of artificial intelligent in trade. *International scientific e-journal АГОС. ONLINE*, 15. DOI: <https://doi.org/10.36074/2663-4139.15.10> [in Ukrainian]

6. Stuzhuk, K.S., Bredihin, V.M. (2021). Artificial intelligence to predict the outflow of customers. *Proceedings of the VI All-Ukrainian scientific-practical conf. applicants for higher education and young scientists Prospects for the development of territories: theory and practice*, 18–19 Nov 2021. [in Ukrainian]

Рецензент: д-р техн. наук, проф. А.Л. Литвинов, Харківський національний університет міського господарства, Україна.

Автор: БРЕДІХІН Володимир Михайлович
кандидат технічних наук, доцент, доцент кафедри комп'ютерних наук та інформаційних технологій
Харківський національний університет міського господарства імені О.М. Бекетова
E-mail – bredixinv@gmail.com
ID ORCID: <https://orcid.org/0000-0002-6063-5046>

Автор: СЕНЧУК Тетяна Сергіївна
старший викладач кафедри комп'ютерних наук та інформаційних технологій
Харківський національний університет міського господарства імені О.М. Бекетова
E-mail – tetyana.senchuk@kname.edu.ua
ID ORCID: <https://orcid.org/0000-0003-0751-5149>

Автор: СТУЖУК Катерина Сергіївна
студентка 4 курсу навчально-наукового інституту енергетичної, інформаційної та транспортної інфраструктури
Харківський національний університет міського господарства імені О.М. Бекетова
E-mail – kateryna.stuzhuk@kname.edu.ua

FOCUS ON ARTIFICIAL INTELLIGENCE FOR PREDICTING THE OUTFLOW OF CLIENTS FROM ON-LINE EDUCATION SITES

V. Bredikhin, T. Senchuk, K. Stuzhuk

O.M. Beketov National University of Urban Economy in Kharkiv, Ukraine

The article examines the process of forecasting customer outflows, which is especially important for companies that use a business model based on subscription. It was found that the outflow rate is extremely important for companies with a subscription and transactional business model, which implies regular payments to the company (banks, telecom operators, SaaS-services, etc.). For this purpose, the types, the main reasons for the outflow of customers and the parameters defined to build a predictive model using machine learning algorithms were considered.

The result was the hypothesis of the reasons for the outflow of customers from sites that provide training services based on courses that are presented on-line in the Internet space. To build a model of outflow forecasting, the behavioral characteristics of students, their motivation and the structure of the courses themselves were studied. Based on the collected large array of data, their change was analyzed by a large number of parameters and the relationships between the behavioral characteristics of students, course structures and their passage were identified. A variant of the forecasting model was built, for which the accuracy of its operation was increased and the results were integrated into the customer outflow prediction module. The final list of features included more than 100 parameters, which were divided into 6 blocks. As a result, a predictive model was created using the Weibull distribution, as client behavior can be considered as a kind of survival model. To estimate the probability of customer outflow, based on the considered hypotheses, a recurrent neural network with an LSTM layer was developed, where a negative logarithmic likelihood function was used as a loss function for the Weibull distribution. As a conclusion, it was proposed to introduce a stable proactive educational business, when decisions are made not only on the basis of feelings, but also on the basis of data, comes a clearer and more sound understanding of how to improve the educational product.

Keywords: *customer outflow, business transaction model, Weibull distribution, neural network, machine learning algorithms.*