

МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ЕКОНОМІЦІ

УДК 330.4

DOI: <https://doi.org/10.32782/2224-6282/161-27>**Піскунова О. В.**доктор економічних наук, професор,
Київський національний економічний університет
імені Вадима Гетьмана**Клочко Р. С.**аспірант,
Київський національний економічний університет
імені Вадима ГетьманаORCID: <https://orcid.org/0000-0003-2690-2785>**Piskunova Olena, Klochko Rostyslav**

Kyiv National Economic University named after Vadym Hetman

ПРОГНОЗУВАННЯ ЧАСУ ДО НАСТУПНОГО ЗАМОВЛЕННЯ КЛІЄНТА ІНТЕРНЕТ-МАГАЗИНУ НА ПІДГРУНТІ МЕТОДІВ МАШИННОГО НАВЧАННЯ

У зв'язку зі швидким розвитком Інтернет-торгівлі та посилення конкуренції на роздрібному ринку України компанії змушені шукати нові шляхи для росту власного бізнесу. Можливим варіантом є підвищення ефективності маркетингової діяльності, досягти якої можна за рахунок прогнозування поведінки споживачів та побудови маркетингової політики з огляду на індивідуальні особливості кожного клієнта. Метою дослідження, представленого у статті, було прогнозування поведінки клієнтів Інтернет-магазину, а саме часу до наступного замовлення, на підґрунті методів машинного навчання та порівняльний аналіз ефективності різних алгоритмів моделювання. Реалізовано п'ять алгоритмів класифікації: лінійний дискримінантний аналіз, дерево рішень, «випадковий ліс», метод опорних векторів, k – найближчих сусідів та здійснено порівняльний аналіз їх ефективності. З огляду на особливості поведінки клієнтів, для прогнозування часу до наступного замовлення запропоновано розглядати такі часові проміжки в майбутньому, коли клієнт зробить наступне замовлення: до двох місяців, від двох до шести місяців, від шести до п'ятнадцяти місяців та відсутність наступного замовлення. Прогнозування таких проміжків дає змогу визначити клієнтів, які з більшою ймовірністю зроблять наступну покупку, і зосередити на них рекламні бюджети або будувати стратегію управління клієнтським досвідом: активувати клієнтів, які пішли, пропонувати знижки клієнтам, які збираються піти. Розглянуто особливості оцінювання якості класифікаційних моделей на основі «матриці невідповідності» за показниками Accuracy, F1, Recall та Precision точності прогнозування. Проведене дослідження дало змогу надати перевагу моделі класифікації «випадковий ліс». Для підвищення якості моделювання було застосовано десятикратну перехресну перевірку. Зважена точність F1 У групах «до двох місяців» та «від двох до шести місяців» досягла 62,5% і 64,1% відповідно. Розроблена модель має зменшити вплив людського фактору на процес прийняття рішень у ході побудови маркетингових стратегій та підвищити їх ефективність.

Ключові слова: Інтернет-торгівля, поведінка споживачів, час до наступного замовлення, прогнозування, машинне навчання, класифікація, матриця невідповідності.

PREDICTING TIME BEFORE THE NEXT ORDER IN THE ONLINE STORE, BASED ON MACHINE LEARNING METHODS

Due to the rapid development of e-commerce and increased competition in the retail market of Ukraine, companies are forced to look for new ways to grow their business. One of the options is to optimize business processes, in particular to increase the efficiency of marketing activities. Predicting consumer behavior is one of the most effective methods of optimizing marketing budgets by building processes based on the individual characteristics of each client. The aim of the study was to predict the behavior of online store customers, namely the time before the next order, based on machine learning methods and a comparative analysis of the effectiveness of different modeling algorithms. Five classification algorithms were implemented: linear discriminant analysis, classification and regression trees, random forest, support vector machine, k – nearest neighbors and comparative analysis of their efficiency was performed. Given the peculiarities of customer behavior for forecasting time to the next order, it is proposed to consider the following time intervals in the future when the customer makes the next order: up to two months, two to six months, six to fifteen months, and without order. Predicting such intervals allows us to identify customers who are more likely to make the next purchase and focus our advertising budgets on them, or build a customer experience management strategy: activate customers who have left, offer discounts to customers who are going to leave. Peculiarities of classification models quality assessment on the basis of the "confusion matrix" according to the forecasting accuracy indicators "Accuracy", "F1", "Recall" and "Precision" is considered. The study allowed us to give preference to the model of classification "random forest". A tenfold cross-validation was used to improve the quality of the simulation. The

weighted accuracy of "F1" in the groups "Up to two months" and "two-six months" reached 62.5% and 64.1%, respectively. The developed model should reduce the influence of the human factor on the decision-making process in the construction of marketing strategies.

Keywords: e-commerce, consumer behavior, time before the next order, forecasting, machine learning, classification, confusing matrix.

JEL classification: C13, C38, C51, C63, C65

Постановка проблеми. На сучасному етапі розвитку економіки України Інтернет-технології проникли майже в усі сфери життя суспільства, одним із проявів чого стала широка популярність онлайн-торгівлі. Усе частіше споживачі користуються послугами Інтернет-магазинів, тому що це має безліч переваг: зручність вибору, нижча ціна, економія часу на покупку тощо. Водночас завдяки розвитку цифрових технологій боротьба за клієнтів переміщується у глобальний простір. У найближчий час на додачу до жорсткої внутрішньої конкуренції очікується поява на українському ринку нових міжнародних гравців. Це ставить перед вітчизняними Інтернет-магазинами нові завдання щодо поліпшення умов реалізації продукції. Відповідно, для українського ритейлу зростає важливість інновацій, які сприяють підвищенню ефективності ведення бізнесу онлайн-продавцями. Інноваційний розвиток даних компаній стає основним чинником їхньої конкурентоспроможності.

У більшості онлайн-компаній значну частку витрат становлять видатки на маркетингову діяльність, а проблема їх оптимізації стала серйозним викликом для всіх гравців ринку. Розрізнення клієнтів, для яких очікується висока активність у майбутньому, і клієнтів, для яких прогнозується пасивна поведінка, та застосування індивідуального підходу має суттєво зменшити маркетингові витрати та підвищити їх ефективність.

Таким чином, прогнозування активності споживачів є актуальною проблемою, розв'язання якої дасть значний поштовх до росту конкурентоспроможності українських Інтернет-торговців. Зокрема, важливим завданням під час побудови ефективної маркетингової стратегії є прогнозування часу до наступного замовлення.

Аналіз останніх досліджень і публікацій. Більшість досліджень у сфері моделювання маркетингової діяльності, представлених в українській науковій літературі, спрямовано на застосування загальних показників діяльності компанії, таких як прибуток [1], продажі [2] тощо. У ході різних досліджень розкриваються особливості моделювання та прогнозування ціни реалізації товару [3], а також попит [4] на нього. У більшості зарубіжних наукових досліджень реалізуються моделі машинного навчання або нейронних мереж для оцінки майбутньої поведінки конкретного клієнта [5; 6]. Слід зауважити, що для прогнозування поведінки клієнта, як правило, використовується лише один алгоритм моделювання, окрім того, відсутній алгоритм оцінювання якості побудованих моделей [7]. Водночас важливим є порівняльний аналіз ефективності різних алгоритмів.

Розвиток цифрових технологій та доступність великих даних, які постійно оновлюються, спричинили бум досліджень із застосування інтелектуальних технологій моделювання в Інтернет-торгівлі. Дослідниками реалізовано модель прогнозування споживчого кошика клієнта [8], модель оцінки релевантності поштових

розсилок [9] та алгоритм прогнозування ймовірності до повернення онлайн-замовлення [10].

Одним з основних завдань в Інтернет-торгівлі є прогнозування наміру клієнта до наступного замовлення. Але аналіз напрацювань у даній сфері показав, що передбачення конкретного дня покупки майже не практикується через неможливість передбачення всіх обставин, які впливають на прийняття рішення про здійснення замовлень [11]. Алгоритми машинного навчання зазвичай допомагають виявити найбільш явні тенденції поведінки споживача. Прикладом може слугувати завдання виявлення покупців, які кожного місяця протягом довгого періоду часу здійснюють однорідні замовлення. У деяких працях для визначення таких клієнтів застосовують технології блокчейн [12]. Проте частка таких клієнтів досить мала, що робить зазначений аналіз малоефективним у маркетинговому менеджменті.

Таким чином, аналіз праць як вітчизняних, так і зарубіжних науковців дає змогу зробити висновок про доцільність переходу від традиційних підходів до прогнозування поведінки споживачів на основі їх соціально-демографічних характеристик або результатів опитування та суб'єктивних суджень експертів [13] до моделювання поведінки споживачів на підґрунті аналізу їх минулої купівельної активності за допомогою інтелектуальних технологій моделювання, зокрема методів машинного навчання. Важливою особливістю даного підходу є спроможність побудованих алгоритмів до постійного автоматичного навчання та поліпшення точності прогнозування.

Отже, застосування методів машинного навчання у прогнозуванні часу до наступного замовлення споживачів Інтернет-магазину може бути досить ефективним підходом до аналізу очікуваної активності клієнтів та управлінні маркетинговою діяльністю Інтернет-магазину.

Мета статті полягає у прогнозуванні поведінки клієнтів Інтернет-магазину, а саме часу до наступного замовлення, на підґрунті методів машинного навчання та порівняльному аналізі ефективності різних алгоритмів моделювання.

Виклад основного матеріалу. Для побудови автоматичної моделі прогнозування поведінки споживачів вибрано трикроковий алгоритм: кластеризація – класифікація – прогнозування [14].

На першому етапі алгоритму під час проведення кластеризації аналізується клієнтська база та здійснюється поділ клієнтів на певну кількість кластерів. На другому етапі, на якому здійснюється класифікація клієнтів, систему навчають розпізнавати дані кластери та автоматично відносити клієнтів до певного кластеру. На третьому етапі за допомогою методів машинного навчання модель тренується автоматично прогнозувати час до наступної покупки на основі даних щодо минулої активності клієнта та з урахуванням кластеру, до якого його віднесено. Кластеризація

та класифікація клієнтів, які виконуються на першому та другому етапах алгоритму, розглянуто у праці [15] авторів статті. Представлене дослідження присвячене третьому етапу алгоритму, тобто безпосередньо прогнозуванню часу до наступного замовлення клієнта Інтернет-магазину.

Аналіз, моделювання та прогнозування виконувалися за допомогою програмного засобу R-Studio.

У дослідженні використано дані стосовно продажів Інтернет-магазину [16], які містять інформацію щодо 1 067 371 транзакції купівлі та повернення товарів в Інтернет-магазині за період з 01.09.2009 по 09.12.2011.

Таблиця даних містить такі поля: Invoice – унікальний код операції; StockCode – унікальний код товару; Description – назва товару; Quantity – кількість купленої/поверненої продукції; InvoiceDate – дата здійснення операції; Price – ціна товару; Customer ID – унікальний код покупця; Country – країна – ініціатор операції.

Для прогнозування часу до наступного замовлення початкову вибірку було поділено на «минулу» і «майбутню» активність. За «минулу» активність клієнта вважався період з 01.12.2009 по 31.05.2011. Прогнозувалася активність клієнтів у період з 01.06.2011 по 09.12.2011. За дату останньої покупки було прийнято максимальну дату по кожному клієнту у «минулій» активності. Датою наступної покупки була мінімальна дата по кожному з клієнтів у вибірці «майбутньої» активності. Часом до наступної покупки була прийнята кількість днів між датою наступної та останньої покупок.

Модель прогнозування часу до наступного замовлення базується на алгоритмах машинного навчання з «учителем» (supervised learning). У цих методах дослідник надає алгоритму пари «об'єкт – відповідь», а алгоритм самостійно знаходить шлях отримання відповіді по об'єкту. Найчастіше алгоритм здатен дати відповідь для об'єкта, якого він ніколи не зустрічав у минулому, без жодного впливу людини.

Класифікація – це тип навчання з учителем, присвячений вирішенню наступного завдання: за наявності безлічі об'єктів, які розподілені на класи, необхідно створити алгоритм, здатний віднести довільний об'єкт до відповідного класу. У розробленій моделі такими класами виступають періоди часу у майбутньому, в які клієнт із найбільшою ймовірністю здійснить наступну покупку. Дані періоди вибираються виходячи з бізнес-потреб компанії. Було вибрано такі класи:

- здійснення покупки буде протягом наступних двох місяців;
- здійснення покупки буде протягом наступних шести місяців;
- період від 6 до 15 місяців;
- більше не буде здійснено жодної операції.

Атрибутами моделі прогнозування часу до наступного замовлення було вибрано показники, які найкраще описують купівельну активність.

1. Кількість днів між останніми чотирма замовленнями клієнта (Dif1, Dif2, Dif3). У ході обчислень було знехтувано тими клієнтами, які зробили за весь період менше чотирьох покупок.

2. Показники середньоквадратичного відхилення та середнє значення кількості днів між усіма покупками, які були здійснені клієнтом (StD, Ave).

3. Сегмент клієнта (Client_Type), який характеризує рівень його купівельної активності:

- втрачені клієнти (Lost) – не купують протягом довгого часу;
- новий оптовий покупець (New_wholesales) – високий середній чек і активність, але пройшло небагато часу з першої покупки. Необхідно приділяти зусилля для підвищенні лояльності даних клієнтів до бізнесу;
- клієнти, яких скоро втратимо/або купують лише раз на рік (Recent) – середні показники активності, але пройшло багато часу з останньої покупки. Клієнти, на яких варто звернути увагу і намагатися вплинути на здійснення більш частих операцій;
- активний роздрібний покупець (Act_retail) – висока активність, протягом довгого часу купує, середній рівень чеку. Найбільш цінний і лояльний тип клієнтів для бізнесу;
- нові роздрібні клієнти (New_retail) – висока активність, купує протягом короткого часу, середній рівень чеку. Слід зосередити зусилля на тому, щоб перетворити їх на постійних клієнтів;
- активний оптовий покупець (Ac_wholesales) – високий середній чек і активність, купує протягом тривалого періоду. Найбільш дохідний тип клієнтів. Приклад вхідної інформації подано в табл. 1.

На початковому етапі процесу моделювання та прогнозування інформацію про «минулу» активність випадковим чином було поділено на навчальну та тестову вибірки: 75% – навчальні, 25% – тестові дані.

Найбільш поширеними методами машинного навчання для задач класифікації є такі: дискримінантний аналіз (Linear discriminant analysis, LDA); метод опорних векторів (Support vector machine, SVM); дерево рішень (Classification and regression trees, CART); k – найближчих сусідів (K-nearest neighbors, KNN); випадковий ліс (Random forests, RF).

Перелічені алгоритми реалізовано на навчальній вибірці. Для оцінки якості моделювання використовувалася оцінка Accurasy, яка є відношенням правильно розподілених клієнтів до їх загальної кількості.

У ході моделювання важливо переконатися, що результати моделювання не залежать від конкретного

Таблиця 1

Попередня вибірка для побудови моделей

Client_ID	Час до наступної покупки	Dif1	Dif2	Dif3	Ave	StD	Client_Type
13085	2-6 місяців	384	2	57	148	207	Act_retail
18102	До 2 місяців	2	1	1	1	1	Act_wholesales
18087	6-15 місяців	89	9	5	34	47	New_retail
15413	Без жодної операції	7	26	246	71	82	Lost
16167	Без жодної операції	1	24	179	68	97	Recent
13767	До 2 місяців	2	5	8	5	3	New_wholesales

набору даних. Для вирішення цієї задачі під час впровадження алгоритму класифікації клієнтів було застосовано десятикратну перехресну перевірку. Десятикратний тест передбачає поділ вибірки на десять випадково вибраних наборів (тестові та навчальні зразки) та тестування побудованої моделі на кожній із них окремо.

У табл. 2 наведено характеристики розподілу показника точності Ассигасу: його мінімальне, максимальне, середнє значення та 1, 2 і 3 квартилі. Дані характеристики отримані на навчальній вибірці для кожного з розглянутих методів.

Як бачимо з табл. 2, на етапі навчання найкраще себе показали алгоритми lda (усереднена точність на рівні 54%) та rf (53%). Побудовані моделі було також перевірено на тестовій вибірці. Найкраще тут показав себе алгоритм rf (усереднена точність передбачень на рівні 54%). Далі для прогнозування часу до наступного замовлення застосовуватимемо саме цей алгоритм.

Результати реалізації на тестовій вибірці методу «випадковий ліс» в абсолютних показниках представлено у табл. 3. У таблиці наведено значення кількості клієнтів, віднесених до визначених класів.

Слід зазначити, що оцінка якості моделей за допомогою метрики Ассигасу необхідно використовувати дуже обережно, оскільки вона досить чутлива до екстремальних значень кількості об'єктів, які утворюють певний клас. Наприклад, якщо існує два класи об'єктів, де в першому класі три об'єкти, а в другому – 97. Алгоритму достатньо всім об'єктам присвоїти другий клас для отримання точності в 97%. З огляду на дану проблему, для оцінки фінальної моделі було застосовано підхід «матриця невідповідності» (confusion matrix).

«Матриця невідповідності» будується на основі чотирьох показників:

- TP – правильно розподілений позитивний клас (true positive);
- TN – правильно розподілений негативний клас (true negative);
- FP – неправильно розподілений позитивний клас (false positive);
- FN – неправильно розподілений негативний клас (false negative).

Позитивним класом є той, перевірка якого відбувається. Негативним – усі інші класи. Наприклад, для обчислення точності прогнозування категорії «до двох місяців» необхідно побудувати таку таблицю-матрицю (табл. 4).

На основі «матриці невідповідностей» обчислюються такі критерії якості класифікації:

точність (precision) – це частка правильно спрогнозованих позитивних прикладів серед прикладів, які модель класифікувала як позитивні;

чутливість (recall) – це частка прикладів, класифікованих як позитивні, серед загальної кількості реально-позитивних прикладів;

показник F_1 , який поєднує у собі точність і чутливість в єдину міру. Математично це середнє гармонічне значення точності та чутливості. Дане значення можна розрахувати так:

$$F_1 = \frac{2}{\frac{1}{recall} + \frac{1}{precision}} = 2 * \frac{precision * recall}{precision + recall} = \frac{tp}{tp + \frac{1}{2}(fp + fn)} \quad (1)$$

Таблиця 2

Результати перехресної перевірки на тренувальній вибірці

Accuracy	Min,	1st Qu,	Median	Mean	3rd Qu,	Max,
lda	0,48	0,52	0,54	0,54	0,55	0,59
cart	0,49	0,51	0,52	0,53	0,55	0,57
knn	0,30	0,31	0,33	0,34	0,35	0,43
svm	0,46	0,53	0,53	0,53	0,54	0,59
rf	0,47	0,51	0,53	0,53	0,54	0,59

Таблиця 3

Результати прогнозування за методом «випадковий ліс»

Період наступної покупки		Реальні значення			
		До 2 місяців	2-6 місяців	6-15 місяців	Без жодної операції
Прогнозне значення	До 2 місяців	61	11	5	4
	2-6 місяців	53	126	24	45
	6-15 місяців	0	7	40	47
	Без жодної операції	0	1	10	15

Таблиця 4

Період наступної покупки		Реальні значення			
		до 2 місяців (позитивний)	2-6 місяців	6-15 місяців	без жодної операції
Прогнозне значення	до 2 місяців (позитивний)	TP	FP	FP	FP
	2-6 місяців	FN	TN	TN	TN
	6-15 місяців	FN	TN	TN	TN
	без жодної операції	FN	TN	TN	TN

Таблиця 5

Результати обчислення показників якості моделі

	Precision	Recall	F1-score
До 2 місяців	75.31%	53.51%	62.56%
2-6 місяців	50.81%	86.90%	64.12%
6-15 місяців	42.55%	50.63%	46.24%
Без жодної операції	57.69%	13.51%	21.90%

У табл. 5 подано результати розрахунку критеріїв якості моделі за методом «матриця невідповідності» для всіх наявних класів.

Як можна побачити, не маючи ніякого уявлення про купівельні переваги клієнтів на початку дослідження, ми спромоглися отримати до 64% точних передбачень періоду до наступного замовлення.

Даний алгоритм налаштований на постійне навчання і розвиток. Передбачення майбутньої активності клієнта ставатиме більш точним із кожною наступною здійсненою операцією, оскільки буде можливість відслідковувати періоди між покупками не лише останніх чотирьох операцій, а й усіх здійснених клієнтом замовлень. Але навіть розробленого нами

алгоритму цілком достатньо, щоб уже зараз приймати управлінські рішення.

Висновки. Таким чином, у ході дослідження за допомогою методів машинного навчання було побудовано модель прогнозування часу до наступного замовлення клієнта онлайн-магазину. Було реалізовано п'ять алгоритмів класифікації: лінійний дискримінантний аналіз, дерево рішень, «випадковий ліс», метод опорних векторів, k – найближчих сусідів. За критерієм Accurasy перевагу було надано методу «випадковий ліс» – до 54% точних передбачень.

Для уточнення якості моделювання було побудовано «матрицю невідповідності» та обраховано критерії оцінки точності прогнозування precision, recall, F1. Зважена точність F1 у групах «до двох місяців» та «2–6 місяців» досягла 62,5% і 64,1% відповідно.

Використовуючи здійснений розподіл клієнтів, власники Інтернет-магазину мають змогу: зосереджувати рекламні бюджети лише для тих клієнтів, які з великою ймовірністю зроблять наступну покупку; будувати стратегію управління клієнтським досвідом (активувати клієнтів, які пішли, пропонувати знижки клієнтам, які збираються піти, тощо).

Список використаних джерел:

1. Хорошун В.В., Науменко І.А. Економіко-математичні методи та моделі прогнозування збутової логістики торговельного підприємства. *Причорноморські економічні студії*. 2018. Вип. 28(2). С. 179–183. URL: [http://nbuv.gov.ua/UJRN/bses_2018_28\(2\)_38](http://nbuv.gov.ua/UJRN/bses_2018_28(2)_38) (дата звернення: 01.11.2020).
2. Лозовська Л.І., Чорнорот Я.О. Застосування комплексно-значних моделей для прогнозування поведінки споживача. *Економічний вісник Запорізької державної інженерної академії*. 2016. Вип. 4(2). С. 115–119. URL: [http://nbuv.gov.ua/UJRN/evzdia_2016_4\(2\)_26](http://nbuv.gov.ua/UJRN/evzdia_2016_4(2)_26) (дата звернення: 01.11.2020).
3. Міхайлуца О., Пожуєв А., Тищенко В. Методи інтелектуального аналізу даних та їх застосування у сфері електронної комерції. *Математичне моделювання*. 2020. С. 154–164. DOI: 10.31319/2519-8106.1(42)2020.207020 (дата звернення: 01.11.2020).
4. Голованова М.А., Голованов Д.С. Застосування ланцюга Маркова для визначення місткості ринкових сегментів в умовах цифрової трансформації поведінки покоління споживачів. *Системи управління, навігації та зв'язку*. 2019. Вип. 3. С. 89–98. URL: http://nbuv.gov.ua/UJRN/suntz_2019_3_17 (дата звернення: 01.11.2020).
5. Lang T., Rettenmeier M. (2017) Understanding consumer behavior with recurrent neural networks. In Workshop on Machine Learning Methods for Recommender Systems. URL: <http://userpage.fu-berlin.de/tlang/pub/2017-lang-rettenmeier-mlrec.pdf> (дата звернення: 01.11.2020).
6. Prediction is very hard, especially about conversion / L. Bigon et al. *Predicting user purchases from clickstream data in fashion e-commerce*. arXiv preprint arXiv:1907.00400 (дата звернення: 01.11.2020).
7. Topal Ibrahim. (2019) Estimation of Online Purchasing Intention Using Decision Tree. *Karar Ağacı Kullanarak Çevrimici Satın Alma Niyetinin Tahmini*. № 17. P. 269–280. DOI : 10.11611/yead.542249 (дата звернення: 01.11.2020).
8. Лесна Н.С., Літвінцева А.В. Методи аналізу поведінки покупця Інтернет-магазину для підвищення рівня продажів. *ScienceRise*. 2015. № 12(2). С. 6–10. URL: [http://nbuv.gov.ua/UJRN/texs_2015_12\(2\)_2](http://nbuv.gov.ua/UJRN/texs_2015_12(2)_2) (дата звернення: 01.11.2020).
9. Логін В.В. Підвищення ефективності масової рекламної розсилки. *International scientific journal*. 2016. № 7. С. 89–94. URL: http://nbuv.gov.ua/UJRN/mnj_2016_7_21 (дата звернення: 01.11.2020).
10. Нарушинська О.О., Теслюк В.М., Денисюк П.Ю. Застосування методів машинного навчання для мінімізації повернень товару в системах електронної комерції. *Актуальні проблеми в економіці*. 2017. № 3. С. 342–347. URL: <https://www.researchgate.net/publication/328998595> (дата звернення: 01.11.2020).
11. Gonzalez Munoz, Mario Hedström, Philip Hedström. Predicting Customer Behavior in E-commerce using Machine Learning. (2019). *School of Electrical Engineering and Computer Science (EECS)*. P.10 URL: <http://kth.diva-portal.org/smash/get/diva2:1355056/FULLTEXT01.pdf> (дата звернення: 01.11.2020).
12. Zhang Huibing, Dong Junchao. Prediction of Repeat Customers on E-Commerce Platform Based on Blockchain. (2020). *Wireless Communications and Mobile Computing*. P. 1–15. DOI: 10.1155/2020/8841437 (дата звернення: 01.11.2020).
13. Пічкі К.В. Моделювання поведінки споживачів на ринку послуг. *Вісник Київського національного університету технологій та дизайну. Серія «Економічні науки»*. 2016. № 3. С. 93–97. URL: http://nbuv.gov.ua/UJRN/Vknudt_2016_3_12 (дата звернення: 01.11.2020).
14. Alapati Yaswanth Kumar, Korrapati Sindhu. Combining clustering with classification: a technique to improve classification accuracy. *Lung Cancer*. (2016). *International Journal of Computer Science Engineering (IJCSE)*. Vol. 5. № 06. URL: <http://www.ijcse.net/docs/IJCSE16-05-06-026.pdf> (дата звернення: 01.11.2020).
15. Piskunova O., Klochko R. Classification of e-commerce customers based on Data Science techniques. (2020). URL: <http://ceur-ws.org/Vol-2649/paper2.pdf> (дата звернення: 01.11.2020).
16. Online Retail II Data Set. URL: <https://archive.ics.uci.edu/ml/datasets/Online+Retail+II> (дата звернення: 01.10.2020).

References:

1. Khoroshun V.V. Naumenko I.A. (2018) Ekonomiko -matematychni metody ta modeli prohnozuvannya zbutovoi lohistyky torhovelnoho pidpriemstva. [Economic-mathematical methods and models of prognostication the supply logistics of trade enterprise]. Prychornomorski ekonomichni studii [Black sea economic studies], vol. 28, no. 2, pp. 179-183. Available at: [http://nbuv.gov.ua/UJRN/bses_2018_28\(2\)_38](http://nbuv.gov.ua/UJRN/bses_2018_28(2)_38) (accessed: 01.11.2020)
2. Lozovska L.I. Chornorot Ya.O. (2016) Zastosuvannya kompleksnoznachnykh modelei dlia prohnozuvannya povedinky spozhyvacha. [Application of complex models for forecasting consumer behavior]. Ekonomichnyi visnyk Zaporizkoi Derzhavnoi Inzhenernoi Akademii [Economic journal Zaporizhzhya state engineering academy], vol. 4, no. 2, pp. 115-119. Available at: [http://nbuv.gov.ua/UJRN/evzdia_2016_4\(2\)_26](http://nbuv.gov.ua/UJRN/evzdia_2016_4(2)_26) (accessed: 01.11.2020)
3. Mikhailutsa O., Pozhuiev A., Tyshchenko B. (2020) Metody intelektualnoho analizu danykh ta yikh zastosuvannya u sferi elektronnoi komertsii [Methods of intellectual analysis of data and their application in the field of electronic commerce]. Matematychni modeliuvannya [Mathematical modeling]. vol. 42, no. 2, pp. 154-164. DOI: 10.31319/2519-8106.1(42)2020.207020 (accessed: 01.11.2020)
4. Holovanova M.A. Holovanov D.S. (2019) Zastosuvannya lantsiuha markova dlia vyznachennia mistkosti rynkovykh sehmentiv v umovakh tsyfrovoy transformatsii povedinky pokolin spozhyvachiv [Application of mark's chains for determination of capacity of market segments in the conditions of digital transformation of conduct of generations of consumers]. Systemy upravlinnia navihatsii ta zviazku [Control, Navigation and Communication Systems], vol. 3, pp. 89-98 Available at: http://nbuv.gov.ua/UJRN/suntz_2019_3_17 (accessed: 01.11.2020)
5. Lang T., Rettenmeier M. (2017) Understanding consumer behavior with recurrent neural networks. In Workshop on Machine Learning Methods for Recommender Systems. Available at: <http://userpage.fu-berlin.de/tlang/pub/2017-lang-rettenmeier-mlrec.pdf> (accessed: 01.11.2020)
6. Bigon, L., Cassani, G., Greco, C., Lacasa, L., Pavoni, M., Polonioli, A. and Tagliabue, J., 2019. Prediction is very hard, especially about conversion. Predicting user purchases from clickstream data in fashion e-commerce. arXiv preprint arXiv:1907.00400 (accessed: 01.11.2020)
7. Topal, Ibrahim. (2019) Estimation of Online Purchasing Intention Using Decision Tree. Karar Ağacı Kullanarak Çevrimiçi Satın Alma Niyetinin Tahmini. no. 17, pp. 269-280. DOI: 10.11611/yead.542249 (accessed: 01.11.2020)
8. Liesna N. S. Litvintseva A. V. (2015) Metody analizu povedinky pokuptsia internet mahazynu dlia pidvyshchennia rivnia prodazhiv [Methods of analyzing the behavior of the buyer of the online store to increase sales]. ScienceRise. no. 12, pp. 6-10. Available at: [http://nbuv.gov.ua/UJRN/text_2015_12\(2\)_2](http://nbuv.gov.ua/UJRN/text_2015_12(2)_2) (accessed: 01.11.2020)
9. Lohin V.V. (2016) Pidvyshchennia efektyvnosti masovoi reklamnoi rozsylyky [Improving the efficiency of the mass advertising mailing]. International scientific journal. no. 7, pp. 89-94. Available at: http://nbuv.gov.ua/UJRN/mnj_2016_7_21 (accessed: 01.11.2020)
10. Narushynska O.O. Tesliuk V.M. Denysiuk P.Yu. (2017) Zastosuvannya metodiv mashynnoho navchannia dlia minimizatsii povornen tovaru v systemakh elektronnoi komertsii [Machine learning methods application for minimization of merchandize return in e-commerce]. Aktualni problemy v ekonomitsi [Actual problems of economics], no. 3, pp. 342-347. Available at: <https://www.researchgate.net/publication/328998595> (accessed: 01.11.2020)
11. Gonzalez Munoz, Mario Hedström, Philip Hedström. (2019) Predicting Customer Behavior in E-commerce using Machine Learning. School of Electrical Engineering and Computer Science (EECS). P. 10 Available at: <http://kth.diva-portal.org/smash/get/diva2:1355056/FULLTEXT01.pdf> (accessed: 01.11.2020)
12. Zhang Huibing, Dong Junchao. (2020) Prediction of Repeat Customers on E-Commerce Platform Based on Blockchain. Wireless Communications and Mobile Computing. pp. 1-15. DOI: 10.1155/2020/8841437 (accessed: 01.11.2020)
13. Pichyk K.V. (2016) Modeliuvannya Povedinky Spozhyvachiv Na Rynku Posluh [Modelling the behavior of consumers at services market]. Visnyk Kyivskoho Natsionalnoho Universytetu Tekhnologii Ta Dizainu. Seriya Ekonomichni Nauky [Bulletin of the Kyiv National University of Technologies and Design], no. 3, pp. 93-97. Available at: http://nbuv.gov.ua/UJRN/Vknutd_2016_3_12 (accessed: 01.11.2020)
14. Alapati Yaswanth Kumar, Korrapati Sindhu. (2016) Combining clustering with classification: a technique to improve classification accuracy. Lung Cancer. International Journal of Computer Science Engineering (IJCSE). vol. 5, no. 06. Available at: <http://www.ijcse.net/docs/IJCSE16-05-06-026.pdf> (accessed: 01.11.2020)
15. Piskunova O., Klochko R. (2020) Classification of e-commerce customers based on Data Science techniques. Available at: <http://ceur-ws.org/Vol-2649/paper2.pdf> (accessed: 01.11.2020)
16. Online Retail II Data Set. Available at: <https://archive.ics.uci.edu/ml/datasets/Online+Retail+II> (accessed: 01.11.2020)