

**МЕТОДИ МАШИННОГО НАВЧАННЯ ДЛЯ АНАЛІЗУ
ФІНАНСОВИХ ВИТРАТ У МОБІЛЬНИХ ЗАСТОСУНКАХ**

Уваров К.М.

e-mail: kyrylo.uvarov@nure.ua

Харківський національний університет радіоелектроніки,
кафедра прикладної математики
м. Харків, Україна

This work is devoted to the application of machine learning methods for the analysis of personal financial expenses in mobile applications. Classification and clustering algorithms that enable automated categorization of transactions and identification of spending patterns are considered. The effectiveness of decision trees, random forests, support vector machines, LSTM networks, and k-means clustering is analyzed in the context of processing real-world financial data. A comparative analysis of the methods is presented. The practical significance of the proposed approaches for improving expense tracking accuracy is substantiated.

Зростання кількості мобільних фінансових застосунків та обсягів персональних транзакційних даних зумовлює актуальність автоматизованого аналізу витрат. Традиційні підходи до управління особистими фінансами, засновані на ручному введенні та категоризації операцій, є трудомісткими і схильними до помилок. Методи машинного навчання (МН) дозволяють автоматично обробляти великі масиви фінансових даних, виявляти приховані закономірності та надавати користувачам персоналізовані рекомендації щодо оптимізації бюджету [1].

Мобільні застосунки для відстеження витрат отримують дані з різних джерел: банківські API, фотографії чеків, ручне введення. Неоднорідність і неструктурованість цих даних ускладнюють їх обробку та вимагають застосування комплексу методів попередньої підготовки та інтелектуального аналізу. Серед актуальних задач виділяють: автоматичну категоризацію транзакцій, виявлення аномальних витрат, прогнозування майбутніх витрат та формування бюджетних рекомендацій. Якість вхідних даних безпосередньо впливає на ефективність моделей МН, тому етап їх очищення та нормалізації є критично важливим.

Для класифікації транзакцій за категоріями (харчування, транспорт, розваги, комунальні платежі тощо) застосовуються методи керованого навчання. Древа рішень відзначаються високою інтерпретованістю, що є важливою властивістю у фінансових застосунках, де користувач має розуміти логіку прийнятого рішення. Метод опорних векторів (SVM) ефективний для виявлення аномальних транзакцій завдяки стійкості до викидів у просторі ознак. Логістична регресія також застосовується як базова модель завдяки простоті інтерпретації та низьким обчислювальним витратам. Ансамблевий

метод Random Forest демонструє найвищу точність класифікації серед класичних алгоритмів завдяки агрегації результатів множини дерев рішень та зниженню дисперсії моделі [2].

Важливим аспектом є попередня обробка текстових описів транзакцій. Назви операцій, що надходять від банківських API, нерідко містять скорочення, технічні коди або іноземні символи. Для їх нормалізації застосовуються методи обробки природної мови: TF-IDF-векторизація та word embeddings. Попередня токенізація та лематизація текстових полів підвищують точність класифікаційних моделей на 5–12% порівняно з використанням «сирих» текстових даних [3].

Методи некерованого навчання, зокрема k-means та DBSCAN, використовуються для виявлення прихованих патернів витрат без попередньої розмітки даних [4]. Кластеризація дозволяє ідентифікувати нетипові транзакції, що можуть свідчити про шахрайство або помилкові списання, а також виявляти циклічні фінансові поведінки: регулярні передплати, сезонні піки витрат. Алгоритм DBSCAN є особливо ефективним для виявлення аномалій, оскільки не вимагає задавати кількість кластерів заздалегідь і здатен виокремлювати шумові точки.

Для прогнозування майбутніх витрат застосовуються моделі аналізу часових рядів. Рекурентні нейронні мережі типу LSTM (Long Short-Term Memory) демонструють найвищу точність серед розглянутих підходів при роботі з послідовними фінансовими даними, оскільки здатні враховувати довгострокові залежності між транзакціями. Інтеграція LSTM-моделей у мобільні застосунки потребує їх оптимізації для роботи в умовах обмежених обчислювальних ресурсів, що досягається методами квантизації та використанням легковагових архітектур у рамках підходу TinyML [5].

Порівняльний аналіз розглянутих методів МН наведено в таблиці 1.

Як видно з таблиці 1, методи на основі ансамблів (Random Forest) забезпечують найвищу точність класифікації транзакцій, тоді як LSTM є оптимальним вибором для задач прогнозування часових рядів витрат. Для задач виявлення аномалій без навчальної розмітки перевагу мають методи кластеризації.

Практична реалізація систем МН для аналізу витрат стикається з суттєвими викликами у сфері конфіденційності даних. Фінансова інформація належить до категорії чутливих персональних даних, обробка яких регулюється вимогами GDPR та національного законодавства. Перспективним підходом до вирішення цієї проблеми є федеративне навчання, що дозволяє навчати моделі МН на децентралізованих даних безпосередньо на пристроях користувачів без їх передачі на сервер. Це забезпечує збереження конфіденційності та водночас дозволяє моделі узагальнювати знання з великої кількості користувачів [3].

Таблиця 1 – Порівняння методів машинного навчання для аналізу фінансових витрат

Метод МН	Задача	Точність, %	Переваги	Обмеження
Дерево рішень	Класифікація категорій	82–87	Інтерпретованість, швидке навчання	Схильність до перенавчання
SVM	Бінарна класифікація аномалій	85–90	Стійкість до викидів	Погано масштабується
LSTM	Прогнозування часових рядів	88–93	Врахування послідовності	Потребує великих даних
k-means	Кластеризація патернів	–	Не потребує розмітки	Чутливість до викидів
Random Forest	Класифікація категорій	89–94	Висока точність, стійкість	Висока обчислювальна вартість

Таким чином, методи машинного навчання є ефективним інструментом для автоматизованого аналізу фінансових витрат у мобільних застосунках. Поєднання алгоритмів класифікації, кластеризації та прогнозування часових рядів дозволяє будувати комплексні інтелектуальні системи підтримки фінансових рішень. Перспективними напрямками подальших досліджень є застосування трансформерних архітектур для обробки мультимодальних фінансових даних, а також розвиток методів федеративного навчання для мобільних платформ з урахуванням вимог до енергоефективності та латентності.

Список використаних джерел:

1. Zgurovsky M. Z., Pankratova N. D. System Analysis: Theory and Applications. Springer, 2007. 479 p.
2. McMahan B. et al. Communication-Efficient Learning of Deep Networks from Decentralized Data. Proceedings of AISTATS, 2017. P. 1273–1282.
3. Євсєєва О. Ю., Носков В. І. Методи обробки природної мови у задачах класифікації текстових даних: навч. посіб. Харків: ХНУРЕ, 2019. 144 с.
4. Чала О. В., Назаренко О. П. Інтелектуальний аналіз даних: навч. посіб. Харків: ХНУРЕ, 2020. 172 с.
5. Warden P., Situnayake D. TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers. O'Reilly Media, 2020. 504 p.