

ЙМОВІРНІСНІ МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ТА МОДЕЛЮВАННЯ ФІНАНСОВИХ ЧАСОВИХ РЯДІВ

Кліменко Я.К.

e-mail: yaroslav.klimenko@nure.ua

Науковий керівник – д.т.н., доц. Шафроненко А.Ю.

Харківський національний університет радіоелектроніки, каф. ІНФ
м. Харків, Україна

The work proposes a probabilistic approach to personal financial data analysis and forecasting. The developed full-stack web application implements probability distribution parameters for transaction categorization, time series analysis with stochastic models, and adaptive predictive algorithms for goal achievement estimation. API integration with banking systems ensures automatic data collection, while probabilistic algorithms provide personalized financial analytics with confidence intervals. Experimental results demonstrate the effectiveness of the proposed approach for financial planning and management.

Розробка інтелектуальних систем аналізу фінансових даних є актуальним напрямком досліджень через зростаючу потребу в ефективному управлінні персональними фінансами [1–4]. Ключовим завданням таких систем є не лише облік транзакцій, але й прогнозування майбутніх фінансових станів на основі часових рядів та інтелектуального аналізу даних [2, 5].

Запропонована система базується на впровадженні ймовірнісних параметрів для аналізу фінансових транзакцій. Визначимо множину транзакцій T як:

$$T = \{t(r)\}_{r=1}^m, \quad (1)$$

$$t(r) = (c(r), a(r), d(r), ct(r), s(r)). \quad (2)$$

де $c(r)$ – категорія транзакції, $a(r)$ – сума, $d(r)$ – дата проведення, $ct(r)$ – тип транзакції (дохід/витрата), $s(r)$ – джерело даних (ручне введення/API), $m = \text{card } T$.

Впроваджений вектор ймовірнісних параметрів $P = (p_1, p_2, \dots, p_k)$ представляє розподіл випадкових величин, пов'язаних з фінансовими транзакціями, де:

1. p_1 – ймовірність віднесення транзакції до певної категорії;
2. p_2 – ймовірність виникнення транзакції певного розміру;
3. p_3 – ймовірність виникнення транзакції у певний період;
4. p_4 – ймовірність сезонних коливань транзакцій;
5. p_5 – ймовірність досягнення фінансової цілі у визначений термін;
6. p_6 – ймовірнісна оцінка ризику перевищення витрат над доходами.

Для прогнозування майбутнього балансу впроваджена ймовірнісна модель з параметрами розподілу доходів $I \sim N(\mu I, \sigma_I^2)$, та витрат

$E \sim N(\mu E, \sigma_E^2)$, що дозволяє обчислити не тільки очікуване значення балансу, але й довірчий інтервал:

$$\begin{aligned} B_{t+n} &= B_t + \sum_{i=1}^n I_i - \sum_{i=1}^n E_i, \\ I &\sim N(\mu I, \sigma_I^2), E \sim N(\mu E, \sigma_E^2), \end{aligned} \quad (3)$$

де B_{i+n} – прогнозований баланс через n періодів, B_t – поточний баланс, I_i та E_i – прогнозовані доходи та витрати для періоду i .

Процес прогнозування включає наступні етапи з використанням ймовірнісних параметрів:

1. Кластеризація транзакцій – групування транзакцій за категоріями з оцінкою ймовірності p_1 для коректної класифікації нових транзакцій.
2. Моделювання розподілів – використання параметрів p_2 і p_3 для побудови щільності розподілу сум та частот транзакцій.
3. Оцінка сезонності – врахування параметру p_4 для моделювання сезонних коливань за формулою:

$$S(m) = \mu_S + \sigma_S \sin\left(\frac{2\pi m}{12}\right), \quad (4)$$

де $S(m)$ – сезонний коефіцієнт для місяця m , μ_S – середнє значення, σ_S – амплітуда коливань.

4. Розрахунок трендів – оцінка параметру тренду з урахуванням стохастичної складової:

$$T(n) = \alpha n + \varepsilon, \quad (5)$$

де $T(n)$ – значення тренду через n періодів, α – коефіцієнт нахилу, $\varepsilon \sim N(0, \sigma_T^2)$ – випадкова складова.

Ймовірність досягнення фінансової цілі G до моменту часу n визначається як:

$$P(T_G \leq n) = P(B_{t+n} \geq G) = \int_G^\infty f_{B_{t+n}}(x) dx, \quad (5)$$

де $f_{B_{t+n}}(x)$ – функція щільності ймовірності випадкової величини B_{t+n}

Для оцінки часу досягнення фінансової цілі G запропоновано алгоритм, що враховує ймовірнісний характер фінансових потоків:

$$T_G = \min\{k: P(B_{t+k} \geq G) \geq p_{threshold}\}, \quad (6)$$

де T_G – прогнозований час досягнення цілі, G – цільова сума, а $p_{threshold}$ – порогове значення ймовірності (наприклад, 0.8)

Оцінка ризику перевищення витрат над доходами реалізується через параметр p_6 за формулою:

$$Risk = P(E > I) = \int_{-\infty}^0 f_{I-E}(x) dx, \quad (7)$$

де $f_{I-E}(x)$ – функція щільності ймовірності різниці між доходами та витратами.

На рисунку 1 наведено графік прогнозування балансу та часу досягнення цілі для одного з тестових наборів даних.

Результати експериментальної оцінки точності прогнозування для різних обсягів історичних даних представлені в таблиці 1.

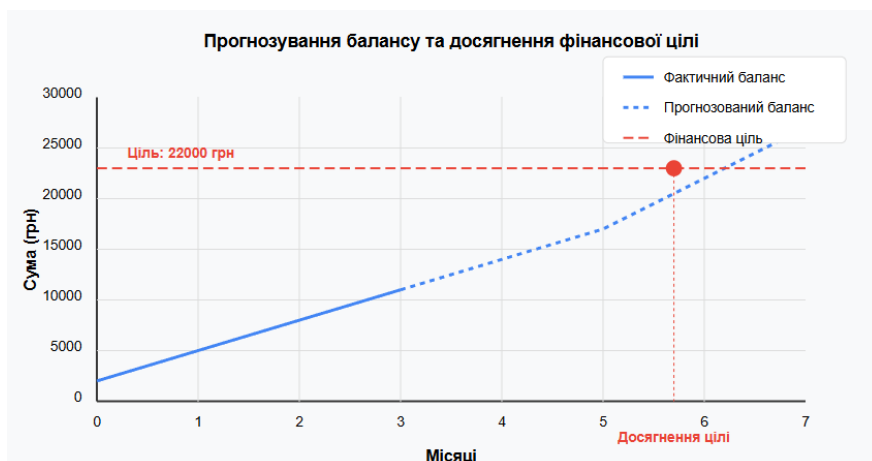


Рисунок 1 – графік прогнозування балансу та часу досягнення цілі

Таблиця 1 – Середня похибка прогнозування балансу (%)

Ймовірнісна модель	1 місяць	3 місяці	6 місяців	12 місяців
Нормальний розподіл	5,1	9,3	15,4	22,6
Гамма-розподіл	4,8	8,7	14,2	20,9
Емпіричний розподіл	3,9	8,1	13,5	19,3
Мультимодальний	3,4	6,8	11,2	18,5

Впровадження ймовірнісних параметрів у систему персонального фінансового менеджменту дозволяє значно підвищити точність прогнозування фінансових показників та ефективність досягнення фінансових цілей. Інтеграція з банківськими API забезпечує більш повний та актуальний набір даних для аналізу, а використання ймовірнісних методів дозволяє адаптувати прогностичні моделі до індивідуальних патернів фінансової поведінки користувача, надати оцінку невизначеності прогнозу.

Список використаних джерел:

1. Hassani H., Silva E.S. Forecasting with Big Data: A Review. *Annals of Data Science*. 2015. Vol. 2, No 1. P. 5–19.
2. Carmona P., Climent F., Momparler A. Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*. 2019. Vol. 61. P. 304–323.
3. Khashei M., Bijari M. An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications*. 2010. Vol. 37, Issue 1. P. 479–489.
4. Ohlhorst F. *Big Data Analytics: Turning Big Data into Big Money*. Wiley, 2012. 176 p.
5. Amirthalingam G., David J.M. A Novel Approach for Financial Forecasting using Deep Learning with Time Series Analysis. *International Journal of Recent Technology and Engineering*. 2019. Vol. 8, Issue 2. P. 3627–3631.