

УДК 004.9

DOI <https://doi.org/10.32689/maup.it.2024.1.3>

**Надія БОЛЮБАШ**

кандидат педагогічних наук, доцент,  
доцент кафедри інтелектуальних інформаційних систем,  
Чорноморський національний університет імені Петра Могили, Nadiya.Bolubash@chmnu.edu.ua  
ORCID: 0000-0002-2274-2422

**Олег ЖЕЛТОБРЮХОВ**

магістрант, кафедра інтелектуальних інформаційних систем,  
Чорноморський національний університет імені Петра Могили, oleg30902228@gmail.com  
ORCID: 0009-0006-4254-8877

**ЧАТ-БОТ ДЛЯ НАДАННЯ РЕКОМЕНДАЦІЙ ІЗ ПЕРЕГЛЯДУ ВІДЕОФІЛЬМІВ  
НА ОСНОВІ МАТРИЧНИХ ФАКТОРИЗАЦІЙНИХ МОДЕЛЕЙ**

**Анотація.** У статті досліджено основні моделі прогнозування реакцій користувача у рекомендаційних системах, засновані на методах матричної факторизації. Обґрунтовано вибір матричної факторизаційної моделі та розглянуто підходи до забезпечення гнучкості взаємодії рекомендаційної системи і користувача шляхом використання чат-бота, впровадженого у вебзастосунок. **Метою статті** є дослідження ефективності застосування чат-бота, орієнтованого на індивідуальні потреби користувача, у рекомендаційній системі з надання рекомендацій по перегляду відеоконтенту на основі матричних факторизаційних моделей. **Методи дослідження.** Використано методи розробки вебзастосунків та інтелектуальних чат-ботів, методи матричної факторизації з використанням методу декомпозиції сигулярного значення SVD, методи машинного навчання, методи обробки й розпізнавання природної мови та методи оптимізації роботи рекомендаційної системи, що базуються на оцінці точності прогнозу та рівня задоволеності користувача спілкуванням із чат-ботом. **Наукова новизна дослідження** полягає у виявленні методів і підходів, спрямованих на покращення отримання користувачами персоналізованих рекомендацій по перегляду відеоконтенту відповідно до їх інтересів та уподобань шляхом застосування чат-бота та моделі прогнозування реакцій користувача на основі методів матричної факторизації. **Висновки.** Накопичення великих обсягів цифрової відеоінформації різних форматів вимагає покращення механізмів надання рекомендацій та підвищення точності прогнозу стосовно уподобань користувачів. Дослідження моделей матричної факторизації MF, машини факторизації FM та машини факторизації з урахуванням поля FFM дозволило установити, що кращі показники точності прогнозу має модель машини факторизації з урахуванням поля FFM: MAE=0,86, MSE=1,65, RMSE=1,28. Для забезпечення гнучкості взаємодії користувача з рекомендаційною системою, розробленою на основі моделі FFM, виявлено доцільність її інтеграції з чат-ботом, впровадженим у вебзастосунок. Дослідження якості створеної моделі обробки природної мови показало високу точність розпізнавання намірів користувача при спілкуванні з чат-ботом – 99,17%. Виявлення рівня задоволеності користувачів спілкуванням з чат-ботом та отриманими рекомендаціями дозволило установити, що задоволеність користувачів становила 86,7%. Що свідчить про високий рівень оцінки ефективності взаємодії користувачів з чат-ботом та високу точність системи стосовно прогнозу намірів користувачів з перегляду відеофільмів.

**Ключові слова:** рекомендаційна система, чат-бот, матрична факторизація, метод декомпозиції сигулярного значення, машинне навчання, машина факторизації з урахуванням поля.

**Nadiia BOLIUBASH, Oleh ZHELTOBRIUKHOV. A CHAT-BOT FOR PROVIDING RECOMMENDATIONS FOR WATCHING VIDEOS BASED ON MATRIX FACTORIZATION MODELS**

**Abstract.** The article examines the main models of predicting user reactions in recommendation systems based on matrix factorization methods. The choice of the matrix factorization model is justified and the approaches to ensuring the flexibility of interaction between the recommendation system and the user through the use of a chatbot implemented in web applications are considered. **The purpose** of the article is to study the effectiveness of using a chatbot in providing personalized recommendations for viewing video content on the basis of a matrix factorization model. **Research methods.** General methods of developing web applications and intelligent chatbots are used, methods of matrix factorization using SVD singular value decomposition method, machine learning methods, natural language processing and recognition methods, and recommendation system optimization methods based on assessment of forecast accuracy, satisfaction level of communication with the chatbot. **The scientific novelty** of the study consists in the identification of methods and approaches aimed at improving users' receipt of personalized recommendations for watching video content in accordance with their interests and preferences by using a chatbot and a model for predicting user reactions based on matrix factorization methods. **Conclusions.** The accumulation of large volumes of digital video information in various formats requires the improvement of mechanisms for providing recommendations and increasing the accuracy of providing predictions regarding user preferences. The research of the matrix factorization models MF, the factorization machine FM, and the field-aware factorization machine FFM made it possible to establish that the model of the field-aware factorization machine FFM had the best indicators of forecast accuracy: MAE=0,86, MSE=1,65, RMSE=1,28. To ensure the flexibility of user interaction with the recommendation system developed on the basis of the FFM model, the expediency of its integration with a chatbot implemented in the web application was found. The research of the quality of the created natural language processing model showed a high accuracy of recognizing the user's

intentions when communicating with the chatbot - 99.17%. Detection of the level of user satisfaction with communication with the chatbot and received recommendations made it possible to establish that user satisfaction was 86.6%. Which indicates a high level of assessment of the effectiveness of user interaction with the chatbot and the high accuracy of the system in terms of predicting users' intentions to watch videos.

**Key words:** recommendation system, chatbot, matrix factorization, singular value decomposition method, machine learning, field-aware factorization machine.

**Постановка проблеми та аналіз останніх досліджень і публікацій.** Накопичення великих обсягів цифрової відеоінформації різних форматів в умовах стрімкого розвитку інформаційного суспільства ускладнює пошук відеоконтенту, який відповідає індивідуальним уподобанням та інтересам користувача [11]. Для вирішення цієї проблеми сервіси, що надають доступ до відео різних типів та жанрів, використовують вбудовані алгоритми фільтрації інформації для надання персоналізованих рекомендацій [4; 12; 17]. Використання рекомендаційних систем дозволяє формувати прогнози стосовно продуктів та послуг, які оптимально будуть відповідати потребам користувачів.

Проте переважна більшість сервісів для перегляду відеоконтенту використовує рекомендаційні системи для вирішення задач комерційного маркетингу і не забезпечує у повній мірі взаємодію між користувачем та системою. Застосування чат-боту дозволяє будувати рекомендації з використанням підходів, які базуються на спілкуванні з користувачем для отримання необхідної інформації про його уподобання та передачі її до рекомендаційної системи для фільтрації відеоконтенту відповідно до його потреб і інтересів [27]. Такий підхід сприяє покращенню формування індивідуальних рекомендацій, дозволяє вирішити проблему холодного старту, спрощує отримання додаткової інформації, забезпечуючи її аналіз у динамічному режимі. Багато застосунків, таких як підтримка клієнтів, інтерактивні платформи обміну повідомленнями, віртуальні помічники та пошук інформації, створені за допомогою чат-ботів [24]. Проведений аналіз досліджень науковців показав, що створення рекомендаційних систем із застосуванням чат-бота у різних предметних сферах дозволяє отримувати результати з високою точністю прогнозу стосовно потреб користувача при наданні рекомендацій та реалізує адаптивні стратегії [7; 9; 20; 27].

У процесі розробки рекомендаційної системи важливим є вибір методів, спрямованих на відбір відеоконтенту з метою забезпечення оптимального прогнозу [15]. Аналіз існуючих підходів до формування прогнозу у рекомендаційних системах дозволив виявити методи колаборативної, контентної та гібридної фільтрації [10; 22]. Контентна фільтрація будує рекомендації, базуючись на інформації про поведінку користувача та його потреби й уподобання. Колаборативна фільтрація будує персональні рекомендації, базуючись на моделі поведінки користувача на основі попередньо зібраної інформації про поведінку інших користувачів із схожими вподобаннями або характеристиками. Серед алгоритмів колаборативної фільтрації виділяють алгоритми, що базуються на даних користувачів та алгоритми, що базуються на даних елементів. Гібридна фільтрація поєднує сильні сторони контентної та колаборативної фільтрації [2].

Підходи, засновані на моделях, при створенні прогнозу рейтингу елементів відеоконтенту використовують накопичену інформацію для формування моделі та її подальшого навчання. Результатом навчання моделі є функція, яка дозволяє отримати рекомендаційний прогноз стосовно потреб та інтересів користувача, виходячи із уже наявних у системі оцінок. При отриманні нових даних модель модернізується для підвищення точності рекомендацій.

Останнім часом для прогнозування реакцій користувача набули широкого поширення моделі на основі матричної факторизації, які мають багато модифікацій: базова матрична факторизація (англ. Matrix factorization, MF), машина факторизації (англ. Factorization Machines, FM), нейронна мережа на основі факторизації (англ. Deep Factorization Machines, DeepFM) [29]. Більш ефективними серед них є моделі, які враховують додаткову інформацію про приховані зв'язки між характеристиками користувачів і елементів та пов'язаними з ними полями: матрична факторизація з урахуванням поля (англ. Field-Aware Matrix Factorization, FMF), тензорна факторизація парної взаємодії (англ. Pairwise Interaction Tensor Factorization, PITF), машина факторизації з урахуванням поля (англ. Field-Aware Factorization Machine, FFM) [13; 30; 31].

Проте ефективність методів матричної факторизації при наданні рекомендацій для перегляду відеоконтенту з використанням чат-бота, орієнтованого на індивідуальні потреби користувача, є не достатньо дослідженою й потребує подальшого опрацювання. Це обумовило **мету статті**, яка полягає у дослідженні ефективності застосування чат-бота у рекомендаційній системі з надання рекомендацій по перегляду відеоконтенту на основі матричних факторизаційних моделей.

**Виклад основного матеріалу.** Рекомендаційна система будує рейтинг елементів відеоконтенту на основі аналізу великого набору даних, що містить інформацію про характеристики елементів і користувачів та уподобання користувачів при виборі елементів.

У результаті проведеного дослідження встановлено, що рекомендаційна система на основі матричної факторизації базується на створенні та навчанні моделі шляхом аналізу даних і наступному використанні навченої моделі для надання користувачам найкращих пропозицій щодо продуктів і послуг. У рекомендаційних системах відеоконтенту необхідно визначити матрицю рейтингів  $R$ , кожен рядок якої представляє користувача, а кожен стовпець – елемент відеоконтенту:

$$R = \begin{pmatrix} r_{11} & \dots & r_{1n} \\ \dots & \dots & \dots \\ r_{m1} & \dots & r_{mn} \end{pmatrix}, \quad (1)$$

де  $r_{ij}$  – оцінка  $j$ -го елемента  $i$ -м користувачем (його рейтинг),  $m$  – кількість користувачів,  $n$  – кількість елементів контенту.

Матриця  $R$  є досить розрідженою, оскільки вона буде містити багато елементів, не оцінених користувачами. Тому у моделі матричної факторизації MF передбачено декомпозицію вихідної матриці на добуток двох матриць меншого рангу (рис. 1).

Для зменшення розмірності матриці рейтингів  $R$  у матричній факторизації використовують низку методів: метод головних компонент (англ. Principal Component Analysis, PCA), невід’ємне розкладання матриці (англ. Non-Negative Matrix Factorization, NMF/NNMF), метод декомпозиції сингулярного значення (англ. Singular Value Decomposition, SVD) і його модифікації Funk-SVD, SVD++, Asymmetric SVD, timeSVD++ [3; 29]. Аналіз досліджень застосування різних методів матричної факторизації показав, що в області відеоконтенту найкращий час навчання мають методи, реалізовані на основі декомпозиції сингулярного значення SVD [23].

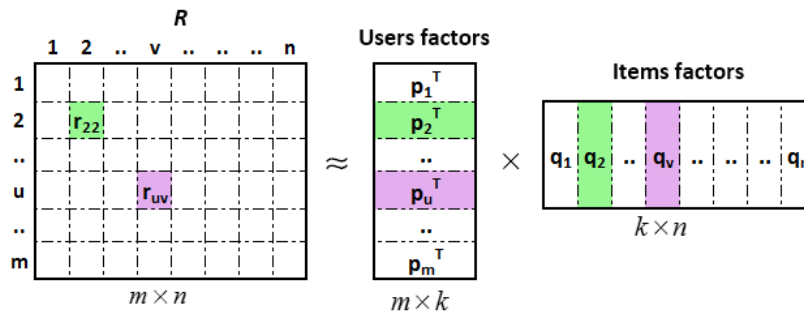


Рис. 1. Схема факторизації матриці рейтингів

Метод SVD передбачає формування матриці  $R'$ , у якій відсутні значення матриці рейтингів  $R$  заповнені з використанням глобального середнього або середнього значення користувача чи елемента [29]. Теорема про сингулярний розклад матриці  $R'$  розмірністю  $m \times n$  дозволяє представити її у вигляді добутку трьох матриць наступним чином:

$$R' = U \cdot S \cdot V^T \quad (2)$$

де  $U$  та  $V$  є унітарними матрицями розмірністю  $m \times m$  та  $n \times n$  відповідно, а матриця  $S$  є діагональною матрицею розмірністю  $m \times n$  із діагональними елементами, рівними сингулярним значенням матриці  $R$  у порядку їх спадання.

У рекомендаційних системах використовують урізане сингулярне розкладання, коли з усіх діагональних елементів матриці  $S$  залишають тільки  $k$  перших елементів, а інші приймають рівними 0. Тоді у матриці  $U$  залишаться тільки  $k$  стовпців, у матриці  $V$  – тільки  $k$  рядків, а матриця  $S$  буде квадратною матрицею розмірністю  $k \times k$ . Щоб виконати декомпозицію, необхідно обрати  $k$  сингулярних значень для складання діагональної матриці  $S_k$  і знайти відповідні рядки та стовпці цих  $k$  сингулярних значень в матрицях  $U$  та  $V$  як  $U_k$  і  $V_k$ . Нова матриця  $R_k^*$  є низькоранговим наближенням матриці  $R$ :

$$R_k^* = U_k \cdot S_k \cdot V_k^T \quad (3)$$

Позначивши  $P = (U_k \cdot S_k)^T$  та  $Q = V_k^T$ , маємо декомпозицію матриці рейтингів у вигляді добутку двох матриць меншої розмірності:

$$R \approx P^T \cdot Q \quad (4)$$

Ранг факторизації  $k$  є величиною низькорангового наближення  $R$ , яку визначають, виходячи з розміру сингулярних значень матриці  $S$ . З цією метою можуть бути використані різні методи [29]: евристичні – відношення суми перших  $k$  квадратів діагональних елементів матриці  $S$  до суми квадратів усіх її діагональних елементів повинно бути більшим за деяке порогове значення (зазвичай рівне 0,95); емпіричні – сума перших  $k$  діагональних елементів матриці  $S$  повинна бути більшою від суми інших сингулярних значень у декілька раз (наприклад, у 10).

Близькість елементів відеоконтенту, представлених у матриці рейтингів, можна визначати з використанням різних мір близькості: відстані Евкліда, квадрату відстані Евкліда, Манхетенської відстані та інших [1, с. 65-66]. У розрідженій матриці рейтингів  $R$  для визначення мір близькості між двома елементами доцільно використовувати косинус подібності:

$$\cos(R_i, R_j) = \frac{R_i \cdot R_j}{\|R_i\| \cdot \|R_j\|}, \tag{5}$$

де  $R_i$  та  $R_j$  є векторами рейтингів  $i$ -го та  $j$ -го елементів, представлених у матриці рейтингів  $R$  у  $i$ -му та  $j$ -му стовпцях, а  $\|R_i\|$  та  $\|R_j\|$  є нормами цих векторів. Аналогічно може бути розрахована близькість двох користувачів, використовуючи їх вектори, представлені у відповідних рядках матриці рейтингів. Під час розрахунку мір близькості у факторизованій матриці рейтингів  $R_k^*$  враховують тільки ті значення, які не є пустими.

Обчислювальну складність традиційного методу SVD можна зменшити, базуючись на навчанні. Такий підхід запропоновано у методах Funk-SVD, SVD++, Asymmetric SVD, timeSVD, які використовують базову ідею методу SVD: матриця рейтингу розкладається на дві матриці нижчого рангу  $Q$  і  $P$ , добуток яких приблизно рівний матриці рейтингів відповідно до формули 4. Отримана наближена матриця рейтингів для елементів з відомими рейтингами буде містити приблизно рівні їм значення, а для елементів з невідомими рейтингами – розраховані прогнозовані рейтинги. Прогнозований рейтинг  $i$ -го елемента відеоконтенту для  $u$ -го користувача розраховують за формулою:

$$r_{ui} = \sum_f p_{uf}^T q_{if}, \tag{6}$$

де  $p_{uf}$  та  $q_{if}$  є відповідними елементами матриць  $Q$  і  $P$ .

Під час навчання моделі матриці  $Q$  і  $P$  знаходять таким чином, щоб мінімізувати функцію втрат – відхилення між відомими рейтингами та їх прогнозами шляхом оптимізації цільової функції. Ранг факторизації є параметром, при якому навчена модель дає кращий прогноз.

У рекомендаційних системах із матричною факторизацією MF не можуть бути використані додаткові характеристики користувачів і відеоконтенту. Однак ця інформація може бути врахована в моделі машини факторизації FM. Для нового фільму матрична факторизація не може визначити рейтинг, оскільки його ніхто не дивився. Проте знаючи жанр, акторів, режисера та інші атрибути фільму, машина факторизації може надавати рекомендації для його перегляду.

Машина факторизації дозволяє враховувати при наданні рекомендацій додаткову інформацію про елементи шляхом комбінації регресії та матричної факторизації. В моделі FM дані для навчання необхідно структурувати, сформувавши вектори ознак  $x$ , перша частина яких кодує користувача, друга – елементи відеоконтенту, наступні – додаткові характеристики користувачів та елементів, поставивши їм у відповідність спостережувані рейтинги  $y$ . Формування навчаючої множини відбувається з використанням елементів матриці рейтингів  $R$  та додаткових атрибутів елементів і користувачів (рис. 2).

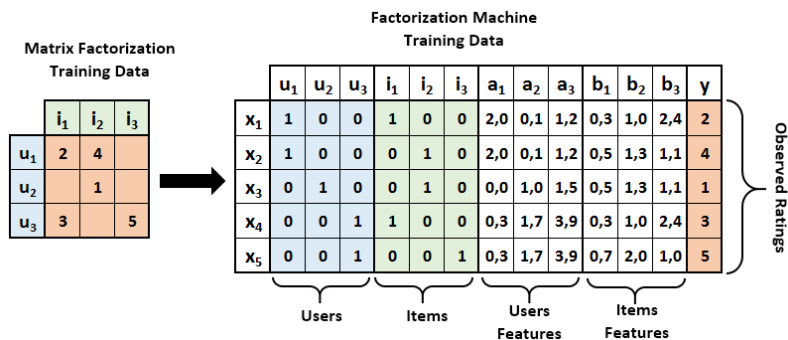


Рис. 2. Структурування даних у моделі машини факторизації FM



Модель FM націлена на моделювання взаємозв'язку між ознаками за допомогою факторизованих параметрів. Для розріджених наборів даних найбільш розповсюдженою є поліноміальна рейгресійна модель другого порядку, яка включає ваги для кожної базової ознаки та для кожної парної комбінації ознак:

$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n w_{ij} x_i x_j, \tag{7}$$

де  $w_0$  – глобальне зміщення,  $n$  – кількість сформованих векторів ознак,  $x_i$  – вектор  $i$ -ї ознаки,  $w_i$  – вага вектора ознаки  $x_i$ ,  $w_{ij}$  – вага для комбінації пари векторів  $i$ -ї та  $j$ -ї ознак,  $\hat{y}(x)$  – вектор прогнозованих рейтингів.

Ранг матриці  $W$  розмірністю  $n \times n$ , яка містить ваги  $w_{ij}$ , знижують шляхом застосування матричної факторизації:

$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j, \tag{8}$$

де  $w_{ij} \approx \langle v_i, v_j \rangle = \sum_{f=1}^k v_{if} v_{jf}$  – скалярний добуток двох векторів розміром  $k < n$ ,  $k$  – ранг факторизації.

Тоді рівняння регресії буде мати вигляд:

$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n x_i x_j \sum_{f=1}^k v_{if} v_{jf}. \tag{9}$$

У процесі навчання моделі, яке здійснюють шляхом мінімізації функції втрат, визначають глобальне зміщення  $w_0$ , ваги векторів ознак  $w_i$  та факторизовані ваги  $v_{if}$ . Навчена модель дає можливість отримувати прогнозований рейтинг із врахуванням додаткових характеристик користувача та елементів відеоконтенту.

Модель машини факторизації з урахуванням поля FFM є розширенням машини факторизації шляхом додавання при формуванні рекомендацій концепції «поля»: схожі характеристики відносять до одного поля [25; 30]. Це дозволяє вирішити проблему моделі MF, яка полягає у тому, що приховані фактори, спільні для ознак, які інтуїтивно представляють різні категорії інформації, можуть погано узагальнювати кореляцію. Наприклад, може бути прихована взаємодія між атрибутами «жанр фільму» та «вік користувача», яка буде проявлятися у тому, що користувачі, які відносяться до різних вікових груп, надають перевагу різним жанрам фільмів. Для фіксації цієї прихованої взаємодії в моделі необхідно виділити поля в атрибутах «жанр фільму» та «вік користувача».

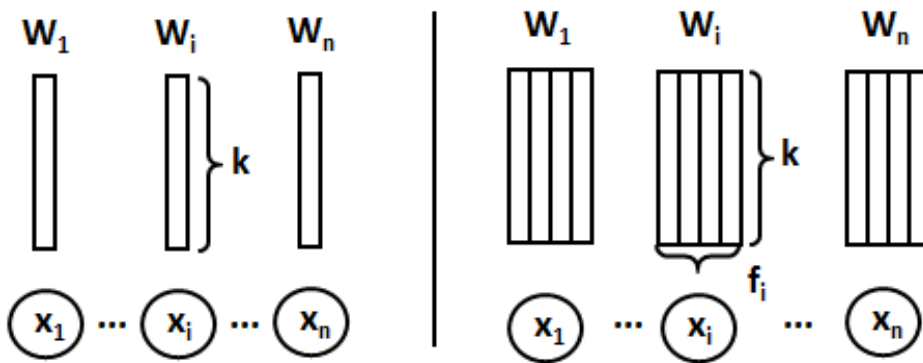
На відміну від стандартної машини факторизації FM, де кожній ознаці відповідає тільки один прихований вектор  $v_i$ , у моделі машини факторизації з урахуванням поля FFM для кожної ознаки створюється стільки векторів, скільки виділено полів. Рівняння поліноміальної регресії другого порядку моделі FFM має вигляд:

$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_{if(i)}, v_{jf(i)} \rangle x_i x_j, \tag{10}$$

де  $v_{if(i)}$  – вага для комбінації ознаки  $x_i$  з полем  $f(i)$  ознаки  $x_j$ ,

$v_{jf(i)}$  – вага для комбінації ознаки  $x_j$  з полем  $f(i)$  ознаки  $x_i$ .

Різницю між моделями FM та FFM візуально відображено на рисунку 3.



а) модель FM б) модель FFM  
**Рис. 3. Порівняння моделей машини факторизації FM та машини факторизації з урахуванням поля FFM**

У моделі машини факторизації вектору кожної ознаки  $x_i$  відповідає вектор її ваг з усіма іншими ознаками  $W_i$ . А в моделі машини факторизації з урахуванням полів вектору кожної ознаки  $x_i$  відповідає матриця ваг  $W_{f_i}$ , яка має  $f_i$  векторів ваг кожного поля ознаки  $x_i$  з полями усіх інших ознак.

Навчання моделі FFM здійснюють шляхом мінімізації функції втрат, визначаючи глобальне зміщення  $w_0$ , ваги векторів ознак  $w_i$  та факторизовані ваги векторів  $v_{f(i)}$  із урахуванням виділених полів. Навчена модель дає можливість отримувати прогнозований рейтинг, у якому враховані приховані зв'язки між різними значеннями характеристик користувачів та відеофільмів.

Першим етапом при створенні рекомендаційної системи було формування набору даних та його попередня обробка. Для побудови рекомендаційної системи було досліджено описані вище моделі на наборі даних у вигляді файлів формату CSV із вебсайту MovieLens, який містить оцінку 58 000 фільмів 280 000 користувачами з січня 1995 року по вересень 2021 року [19]. У наборі даних представлено  $27 \cdot 10^9$  оцінок фільмів у 5-бальній шкалі, характеристики користувачів – «вік» і «стать» та характеристики фільмів – «жанр» і «назва».

Назва фільму у наборі даних була об'єднана з датою виходу фільму у прокат, тому на етапі попередньої обробки даних ці атрибути було розділено з виділенням окремих характеристик кожного фільму: «назва» та «рік виходу». Попередня обробка даних включала також видалення порожніх рядків і невикористаних стовпців та фільтрування даних на основі критеріїв. Для відсіювання маловідомих фільмів було обрано фільми, які оцінили більше, ніж 15 користувачів. Після цих етапів остаточною кількістю рейтингових даних становила 3 0872 062 рядків із 26 897 455 унікальною назвою фільмів і 274 952 унікальним користувачем.

На другому етапі здійснювалося дослідження моделей на основі матричної факторизації та вибір моделі, точність прогнозу якої є найвищою. Створення та навчання моделей матричної факторизації MF і машини факторизації з урахуванням поля FFM здійснювалося з використанням бібліотеки машинного навчання ML.NET (C#) [18]. Модель машини факторизації FM створено та навчено за допомогою бібліотеки Pytorch-Accelerate [28]. У моделі FM при прогнозуванні рейтингів було враховано додаткову інформацію шляхом включення характеристик користувача: «вік» і «стать» та характеристик відеофільмів: «жанр» і «рік виходу». У моделі FFM досліджено ефективність прогнозування з урахуванням прихованої інформації, пов'язаної з полями, які були виділені у характеристиках «вік» і «стать» користувача та «жанр» фільму: 2 статі, 5 вікових груп, 18 жанрів.

Процес навчання моделей, націлених на мінімізацію відхилення між відомими рейтингами та їх прогнозами, реалізовано з використанням методу стохастичного градієнтного спуску шляхом мінімізації функції втрат [8; 26]. Для моделей MF та FM функція втрат визначалася з використанням методу найменших квадратів як сума квадратів помилок для усіх відомих рейтингів і включала фактори регуляризації для уникнення перенавчання рекомендаційної системи. Для моделі FFM було використано логістичну функцію втрат [6; 16].

Під час навчання моделей до навчаючої множини відбиралося 80% об'єктів набору даних, до тестової – 20%. З метою оцінки точності прогнозу було використано наступні показники: MAE, MSE, RMSE (табл. 1). Результати оцінки навчених моделей наведено у таблиці 2.

Таблиця 1

**Показники для оцінювання точності прогнозу рейтингу**  
( $y_i$  і  $\hat{y}_i$  – прогнозоване та фактичне значення рейтингу)

Показник	Опис	Формула
MAE	Середня абсолютна похибка (Mean Absolute Error)	$MAE = \frac{1}{n} \sum_{i=1}^n  y_i - \hat{y}_i $
MSE	Середньоквадратична похибка (Mean Squared Error)	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
RMSE	Корінь квадратний із середньоквадратичної похибки (Root Mean Squared Error)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$

Проведене дослідження дозволило установити, що кращі показники точності прогнозу мала модель машини факторизації з урахуванням поля FFM: MAE=0,86, MSE=1,65, RMSE=1,28. Тому для створення рекомендаційної системи з надання рекомендацій по перегляду відеофільмів було обрано модель FFM.

Таблиця 2

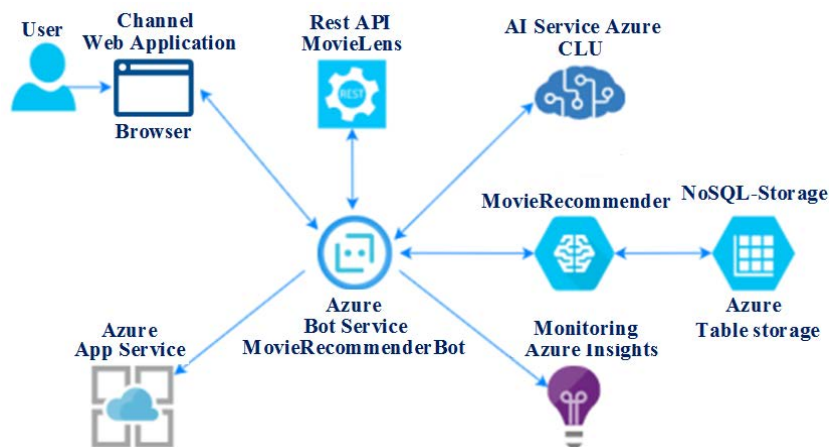
**Результати оцінки точності навчених моделей**

Модель	Показники оцінки		
	MAE	MSE	RMSE
MF	1,02	1,74	1,32
FM	0,89	1,71	1,31
FFM	0,86	1,65	1,28

На наступному етапі створення рекомендаційної системи здійснювалася розробка чат-бота та його інтеграція з навченою моделлю. Чат-бот є програмою або моделлю штучного інтелекту, призначеною для імітації людської розмови за допомогою текстової чи голосової взаємодії. Він може інтерпретувати та розуміти дані користувача і надавати необхідні відповіді у розмовній манері [21]. Спілкування користувачів з чат-ботом при наданні рекомендацій з перегляду відеофільмів забезпечує більшу гнучкість у пошуку відеоконтенту, який відповідає їхнім інтересам.

Створення та управління інтелектуальним чат-ботом для надання рекомендацій реалізовано з використанням Microsoft Bot Framework і Azure Bot Service. Таке поєднання технологій дозволяє використовувати бот як інтегрований компонент у різних каналах спілкування: месенджерах, соціальних мережах, вебзастосунках. Обробка природної мови та розпізнавання намірів користувача здійснювалися із використанням хмарних API Azure Cognitive Services та Conversational Language Understanding (CLU). Сервіс Azure Monitor було застосовано з метою моніторингу й аналізу даних у процесі роботи чат-бота. Для зберігання даних використано NoSQL-сховище Azure Table storage, завдяки чому було реалізовано масштабування у відповідності з потребами. У даному дослідженні чат-бот було інтегровано у середовище вебзастосунку, для розробки якого використано мову розмітки HTML та CSS як засіб стилізації.

Розглянемо архітектуру розробленої рекомендаційної системи (рис. 4). На початку роботи здійснюється виявлення намірів користувача на основі його спілкування з чат-ботом. Користувач взаємодіє з ботом у середовищі вебзастосунку шляхом обміну повідомленнями, для обробки яких використано Azure Cognitive Services та CLU. Отримана інформація стосовно уподобань користувача з перегляду фільмів передається до рекомендаційної системи, яка формує рекомендації та надає результат користувачеві через вебзастосунок. Для інтерактивного відображення рекомендацій у середовищі вебзастосунку використано REST API MovieLens.



**Рис. 4. Архітектура розробленої рекомендаційної системи**

Під час спілкування з чат-ботом на основі аналізу обраних користувачем фільмів та їх оцінки рекомендаційна система накопичує інформацію про його уподобання й інтереси. Це дає можливість покращити точність прогнозу при наданні персоналізованих рекомендацій у подальшому.

Для початку спілкування з чат-ботом із метою виявлення намірів користувача необхідно у вікні браузера відкрити сторінку вебзастосунку та натиснути кнопку «Click me!» (рис. 5).



Рис. 5. Вікно вебзастосунку, інтегрованого з чат-ботом

У розробленій системі чат-бот може спілкуватися з користувачем, задаючи йому питання у вигляді текстових повідомлень, отримувати інформацію про уподобання користувача та надавати на основі цього рекомендації по перегляду фільмів. При створенні бота було навчено та протестовано модель розпізнавання мови користувача з використанням сервісу Conversational Language Understanding із визначенням можливих сценаріїв діалогу, намірів і категорій до кожного сценарію та розробки навчаючих висловлювань для кожного наміру. Оцінка якості створеної моделі обробки природної мови показала високу точність розпізнавання намірів користувача при спілкуванні з чат-ботом – 99,17% (рис. 6).

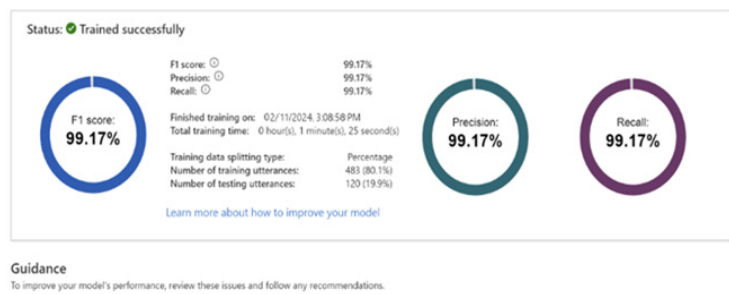


Рис. 6. Оцінка якості та точності моделі розпізнавання мови

При спілкуванні з чат-ботом є дві можливості, які визначають подальший напрямок взаємодії: користувач не вказує назву фільму та користувач вказує назву фільму, який він хоче переглянути. У першому випадку система видає кілька назв популярних фільмів, із яких користувач може обрати один із фільмів як такий, якому він надає перевагу. Передбачена також можливість на основі додаткових питань отримати від користувача бажаний жанр фільму та рік його виходу у прокат. Таким чином реалізується більша гнучкість системи для нових користувачів – обхід проблеми холодного старту. У другому випадку система шукає вказаний фільм та подібні до нього у сховищі даних. Фільми, рекомендовані для перегляду, з коротким описом їх характеристик відображаються у чаті (рис. 7).

Якщо користувач не задоволений наданими результатами, він може повторити запит на отримання інших рекомендацій. Якщо користувач схвально реагує на результат рекомендацій, система припиняє взаємодію з ним, оновлюючи інформацію про рейтингові оцінки користувача. Рекомендовані фільми можна переглянути у будь-якому сервісі, який надає доступ до відеоконтенту.

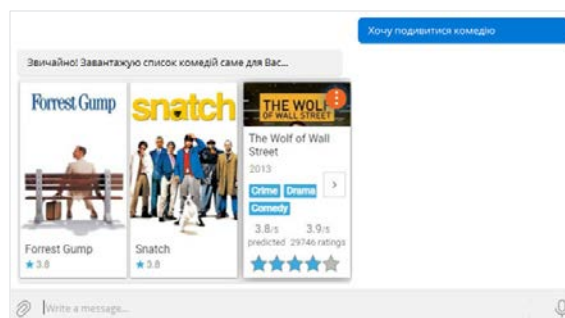


Рис. 7. Приклад рекомендованих для перегляду фільмів



Цілісна оцінка розробленої рекомендаційної системи повинна включати не тільки оцінку точності реалізованих моделей надання рекомендацій, а й якість реалізації розмовних стратегій чат-бота [14]. Це потребує суб'єктивних методів оцінювання, орієнтованих на сприйняття користувачами характеру взаємодії з чат-ботом.

Для дослідження рівня задоволеності користувачів результатами спілкування з чат-ботом та отриманими рекомендаціями було розроблено і впроваджено у вебзастосунок анкету (табл. 3).

Анкета містить 7 питань, згрупованих за шістьма факторами: інформативність (англ. Informative, INF), легкість у використанні (англ. Easy To Use, ETU), передбачувана якість рекомендацій (англ. Perceived Recommendation Quality, PRQ), легкість розуміння (англ. Ease Of Understanding, EOU), довіра (англ. Trust, TR) і сприймана ефективність (англ. Perceived Efficiency, PE) [5].

У кожному питанні анкети користувач може оцінити рейтинг різних факторів взаємодії з чат-ботом на основі визначених вагових коефіцієнтів у балах шляхом вибору наступних варіантів відповідей:

- 1) не згоден (англ. *Disagree*, DS) – 1 бал;
- 2) частково не згоден (англ. *Somewhat Disagree*, SD) – 2 бали;
- 3) нейтрально (англ. *Neutral*, NT) – 3 бали;
- 4) частково згоден (англ. *Somewhat Agree*, SA) – 4 бали;
- 5) абсолютно згоден (англ. *Strongly Agree*, ST) – 5 балів.

Таблиця 3

Питання анкети для оцінювання якості взаємодії з чат-ботом

ID	Фактор	Питання
P1	ETU	Ви можете легко орієнтуватися та взаємодіяти з чат-ботом
P2	ETU	Інструкції та вказівки, які надає система, чіткі та інтуїтивно зрозумілі
P3	EOU	Вам легко зрозуміти рекомендації, які надає система
P4	PE	Ви задоволені швидкістю та чутливістю системи
P5	PRQ	Рекомендовані відео відповідають вашим уподобанням та інтересам
P6	INF	Система надає достатньо інформації про кожен рекомендований фільм
P7	TR	Ви скористаєтеся рекомендаціям системи для майбутнього вибору фільмів

Анкетування проводилося онлайн серед 30 респондентів віком від 17 до 26 років, більшість з яких були студентами. Віковий діапазон містив користувачів, добре знайомих із цифровими технологіями, які часто переглядають відеофільми. За отриманими результатами остаточний бал кожного питання в анкеті  $K_j$  розраховувався за формулою:

$$K_j = \frac{\sum_{i=1}^n k_i^j}{k_{\max}^j}, \quad (11)$$

де  $k_i^j$  – оцінка у балах  $i$ -м опитуваним фактору взаємодії, відображеному у  $j$ -му питанні анкети,  $n$  – кількість опитаних респондентів,

$k_{\max}^j$  – максимальна кількість балів, яку можна отримати за  $j$ -те питання.

Враховуючи, що максимальний ваговий коефіцієнт для кожного питання рівний 5-ти, а кількість респондентів становила 30, для кожного  $j$ -го питання анкети  $k_{\max}^j = 5 \cdot n = 5 \cdot 30 = 150$ .

Рівень задоволеності фактором взаємодії, відображеному у кожному питанні анкети, розраховують у процентах:  $I_j = K_j \cdot 100\%$ . Загальний рівень задоволеності користувачів взаємодією з чат-ботом визначають за формулою:

$$L = \frac{\sum_{j=1}^m I_j}{m}, \quad (12)$$

де  $m=7$  – кількість питань анкети.

У таблиці 4 показано результати проведеного за допомогою розробленої анкети тестування, які показують, що запропонована система рекомендацій на основі чат-бота може забезпечити задовільні результати для користувачів із кінцевим рівнем їх задоволеності 86,6%.

У питаннях P1 і P2 із фактором ETU – легкість у використанні, отримано позитивні результати від користувачів із рівнем задоволення 85,65%. Результати оцінки фактору EOU – легкості у розумінні отриманих рекомендацій, становлять 90%. Це вказує на те, що потік інструкцій під час спілкування з чат-ботом добре інтерпретують користувачі. Вони також задоволені швидкістю та чутливістю чат-бота, який відповідає на запити користувачів за короткий час: рівень задоволення фактору PE – 86%.

Оцінка користувачами фактору PRQ – точності рекомендацій відеофільмів, становила 86%. Повнота отриманої інформації та наміри користуватися ситемою у майбутньому оцінені користувачами в 85,3% і 87,3%. У цілому рівень задоволеності користувачів спілкуванням із чат-ботом та отриманими під час спілкування рекомендаціями становив 86,6%.

Таблиця 4

## Результати опитування щодо задоволення потреб користувачів

Питання анкети		Кількість відповідей на питання анкети					Рівень задоволеності	
Pj	Фактор	DS (1 бал)	SD (2 бали)	NT (3 бали)	SA (4 бали)	ST (5 балів)	K <sub>p</sub> бали	I <sub>p</sub> %
P1	ETU			7	10	13	126	84,0%
P2	ETU			5	9	16	131	87,3%
P3	EOU			2	11	17	135	90,0%
P4	PE		1	2	14	13	129	86,0%
P5	PRQ			4	13	13	129	86,0%
P6	INF		1	2	15	12	128	85,3%
P7	TR			2	15	13	131	87,3%
Остаточний результат, L=								86,6%

Проведене дослідження показало, що реалізація рекомендаційної системи на основі моделі машини факторизації з урахуванням полів та її інтеграція з чат-ботом, впровадженим у вебзастосунок, дозволяє надавати користувачам рекомендації з високим рівнем точності. А їх отримання шляхом спілкування з чат-ботом покращує взаємодію з рекомендаційною системою та робить її адаптованою до індивідуальних потреб користувача. Однак у деяких користувачів можуть виникнути проблеми з отриманням рекомендацій стосовно фільмів, інформація про яких відсутня у системі. Додавання до системи інформації про такі фільми частково допомагає цю проблему вирішити.

**Висновки.** Надання персоналізованих рекомендацій в умовах високих темпів росту обсягів цифрового відеоконтенту вимагає виявлення методів надання рекомендацій з високою точністю прогнозу стосовно намірів та уподобань користувачів та оптимальною гнучкістю їх взаємодії з рекомендаційною системою.

З метою виявлення ефективної моделі надання рекомендацій було досліджено моделі матричної факторизації MF, машини факторизації FM та машини факторизації з урахуванням поля FFM. Навчання моделей здійснювалося на наборі даних, який містив оцінку 280 000 користувачами 58 000 фільмів із використанням методу стохастичного градієнтного спуску. Отримані результати показали, що найвищі показники точності мала модель машини факторизації з урахуванням поля: MAE=0,86, MSE=1,65, RMSE=1,28.

Установлено, що інтеграція навченої моделі FFM з інтелектуальним чат-ботом, у якому реалізовано модель обробки природної мови, дозволяє реалізувати різні стратегії надання рекомендацій, орієнтовані на конкретного користувача, забезпечує гнучкість взаємодії з рекомендаційною системою, вирішує проблему холодного старту, спрощує отримання додаткової інформації. Доступ до каналу спілкування з чат-ботом, яке здійснюється шляхом обміну повідомленнями, надає розроблений вебзастосунок. Оцінка якості створеної та навченої моделі обробки природної мови показала високу точність розпізнавання намірів користувача при спілкуванні з чат-ботом – 99,17%. Дослідження ефективності взаємодії користувачів із чат-ботом, отримане шляхом обробки результатів анкетування, показало високий рівень задоволеності користувачів результатами спілкування з чат-ботом та отриманими рекомендаціями – 86,6%.

Таким чином, проведене дослідження дозволило виявити, що рекомендаційна система на основі моделі матричної факторизації з урахуванням поля FFM, інтегрована з чат-ботом, націлена на надання персоналізованих пропозицій щодо перегляду відеофільмів з високим рівнем точності та високим рівнем задоволеності користувачів результатами спілкування з чат-ботом. У процесі користування рекомендаційною системою накопичується інформація стосовно інтересів та потреб користувача і точність рекомендацій стає вищою. Проте є проблема, пов'язана з додаванням у систему інформації стосовно нових фільмів, яка у режимі реального часу не є вирішеною, оскільки потребує перенавчання моделі з оновленим набором даних. Що обумовлює необхідність подальших розробок у цьому напрямі.

## Список використаних джерел:

1. Болюбаш Н. М. Інтелектуальний аналіз даних. Миколаїв: Вид-во ЧДУ ім. П. Могили, 2023. 320 с. URL: <https://dspace.chmnu.edu.ua/jsptui/handle/123456789/1461>

2. Мелешко Є. В., Семенов С. Г., Хох В. Д. Дослідження методів побудови рекомендаційних систем в мережі Інтернет. *Системи управління, навігації та зв'язку*. Вип. 1(47). 2018. С. 131–136. URL: <https://doi.org/10.26906/SUNZ.2018.1.131>
3. Мелешко Є. В., Хох В. Д., Босько В. В. Дослідження матричних факторизаційних моделей рекомендаційних систем. *Системи управління, навігації та зв'язку*. Вип. 6(58), 2019. С. 58–62. URL: <https://doi.org/10.26906/SUNZ.2019.6.058>
4. Abbas M., Riaz M. U., Rauf A., Khan M. T., Khalid S. Context-aware Youtube recommender system. In *International Conference on Information and Communication Technologies (ICICT)*. Karachi: IEEE, 2017. P. 161–164. URL: <https://doi.org/10.1109/ICICT.2017.8320183>
5. Attalariq M., Baizal Z.K.A. Chatbot-Based Book Recommender System Using Singular Value Decomposition. *Journal of Information System Research*. Vol. 4, No 4. 2023. P. 1293–1301. URL: <https://doi.org/10.47065/josh.v4i4.3817>
6. Chin W. S., Zhuang Y., Juan Y. C., Lin C. J. A Learning-Rate Schedule for Stochastic Gradient Methods for Matrix Factorization. In *Advances in Knowledge Discovery and Data Mining. 19th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2015). Lecture Notes in Computer Science*. Vol. 9077. 2015. Springer Cham. P. 442–445. URL: [https://doi.org/10.1007/978-3-319-18038-0\\_35](https://doi.org/10.1007/978-3-319-18038-0_35)
7. Dalton J., Ajayi V. Main R. Vote Goat: Conversational Movie Recommendation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. Ann Arbor MI USA: ACM, 2018. P. 1285–1288. URL: <https://doi.org/10.1145/3209978.3210168>
8. Duchi J., Hazan E., Singer Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*. Vol 12. 2011. P. 2121–2159. URL: <https://jmlr.org/papers/volume12/duchi11a/duchi11a.pdf>
9. Fajari A. N., Baizal A. Chatbot-based Culinary Tourism Recommender System Using Named Entity Recognition. *Journal Imiah Penelitian dan Pembelajaran Informatika*. Vol. 7, No. 4. 2022. P. 1131–1138. URL: <https://doi.org/10.29100/jipi.v7i4.3210>
10. Falk K. Practical recommender systems. Shelter Island, NY: Manning, 2019. 432 p.
11. Fayyaz Z., Ebrahimian M., Nawara D., Ibrahim A., Kashef R. Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities. *Applied Sciences*. Vol. 10(21). 2020. URL: <https://doi.org/10.3390/app10217748>
12. Gomez-Uribe C.A., Hunt N. The netflix recommender system: algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*. Vol. 6, No 4. 2016. P. 1–19. URL: <https://dl.acm.org/doi/10.1145/2843948>
13. Hong F.X., Zheng X.L., Chen C.C. Latent space regularization for recommender systems. *Information Science*. Vol. 360. 2016. P. 202–216. URL: <https://doi.org/10.1016/j.ins.2016.04.042>
14. Jannach D. Evaluating conversational recommender systems: A landscape of research. *Artificial Intelligence Review*. Vol. 56, No. 3. 2023. P. 2365–2400. URL: <https://doi.org/10.1007/s10462-022-10229-x>
15. Jayalakshmi S., Ganesh N., C'ep R., Senthil Murugan J. (2022). Movie recommender systems: concepts, methods, challenges, and future directions. *Sensors*. Vol. 22(13). URL: <https://doi.org/10.3390/s22134904>
16. Juan Y., Zhuang Y., Chin W.S. Field-aware Factorization Machines for CTR Prediction. In *Proceedings of the 10th ACM Conference on Recommender Systems*. 2016. P. 43–50. URL: <https://doi.org/10.1145/2959100.2959134>
17. Ma S., Zha Z., Wu F. Knowing user better: jointly predicting click-through and playtime for micro-video. In *IEEE International Conference on Multimedia and Expo (ICME)*. 2019. P. 472–477. URL: <https://doi.org/10.1109/ICME.2019.00088>
18. ML.NET Documentation. URL: <https://learn.microsoft.com/en-us/dotnet/machine-learning>
19. MovieLens 25M Dataset. URL: <https://grouplens.org/datasets/movielens/25m>
20. Narducci F., Gemmis M.D., Lops P., Semeraro G. Improving the User Experience with a Conversational Recommender System. In *AI\*IA Advances in Artificial Intelligence*. 2018. P. 528–538. URL: [https://doi.org/10.1007/978-3-030-03840-3\\_39](https://doi.org/10.1007/978-3-030-03840-3_39)
21. Nugraha M., Baizal Z.K.A., Richasdy D. Chatbot-Based Movie Recommender System Using POS Tagging. *Building of Informatics, Technology and Science (BITS)*. 2022. Vol. 4, No. 2. P. 624–630. URL: <https://doi.org/10.47065/bits.v4i2.1908>
22. Padti P. G., Hegde K., Kumar P. Hybrid Movie Recommender System. *International Journal of Research in Engineering, Science and Management*. Vol. 4, no. 7. 2021. P. 311–314.
23. Pujahari A., Sisodia D. S. Model-Based Collaborative Filtering for Recommender Systems: An Empirical Survey. In *2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*. Raipur, India: IEEE, 2020. P. 443–447. URL: <https://doi.org/10.1109/ICPC2T48082.2020.9071454>
24. Singh A., Ramasubramanian K., Shivam S. Building an Enterprise Chatbot: Work with Protected Enterprise Data Using Open Source Frameworks. Berkeley, CA: Apress. 2019. 385 p. URL: <https://doi.org/10.1007/978-1-4842-5034-1>
25. Sun J., Zhang A., Pan J., Flores A. Field-matrixed Factorization Machines for Recommender Systems. In *WWW'21: Proceedings of the Web Conference*. 2021. P. 2828–2837. URL: <https://doi.org/10.1145/3442381.3449930>
26. Taddy M. Stochastic Gradient Descent. In *Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions*. New York: McGraw-Hill. 2019. P. 303–307.
27. Theosaksomo D., Widyantoro D. H. Conversational Recommender System Chatbot Based on Functional Requirement. In *IEEE 13th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*. 2019. P. 154–159. URL: <https://doi.org/10.1109/TSSA48701.2019.8985467>
28. Welcome to pytorch-accelerated's documentation. URL: <https://pytorch-accelerated.readthedocs.io/en/latest>
29. Zhang Y. An Introduction to Matrix factorization and Factorization Machines in Recommendation System, and Beyond. 2022. URL: <https://doi.org/10.48550/arXiv.2203.11026>
30. Zhang Z., Lui Y., Zhang Zh. Field-Aware Matrix Factorization for Recommender Systems. *IEEE Access*. Vol. 6. 2018. P. 45690–45698. URL: <https://doi.org/10.1109/ACCESS.2017.2787741>
31. Zhao X., Li X., Liao L., Song D., Cheung W.K. Crafting a time-aware point-of-interest recommendation via pairwise interaction tensor factorization. In *Knowledge Science, Engineering and Management. KSEM*. 2015. *Lecture Notes in Computer Science*. Vol. 9403. Cham, Switzerland: Springer. P. 458–470. URL: [https://doi.org/10.1007/978-3-319-25159-2\\_41](https://doi.org/10.1007/978-3-319-25159-2_41)