

Трегубчак І.М., Ільїн О.Ю., Коба А.Б.

Державний університет телекомунікацій, Київ

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИБОРУ МАРШРУТІВ З ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО НАВЧАННЯ

Анотація. Інформаційні технології проникають у всі сфери життя людини та діяльності фірм. Не оминула їх вплив і сфера туризму. Туризм – це діяльність у дозвіллі, що включає складні процеси прийняття рішень, саме тому актуальним є розробка рекомендаційних систем, які допоможуть полегшити ці процеси. Саме для цього в роботі була створена методика вибору туристичних маршрутів з використанням методів машинного навчання. В процесі дослідження проаналізована сфера туризму, встановлено особливості рекомендаційних систем. Описані рівні запропонованої методики, вказавши особливості бекенду та фронтенду. Характерною особливістю запропонованої методики є використання в моделі оптимального вибору класифікаторів машинного навчання, таких як дерева рішень, метод опорних векторів та багатошаровий перцептрон. Саме така комбінація методів дозволить обрати найкращий маршрут із запропонованих. Вони використовуються для розрізнення конкретних пунктів призначення в кожному наборі даних. Щоб зробити комплексну модель придатною для використання та інтерпретувати її результати для туриста, моделі дерева рішень перетворюються на правила прийняття рішень, а потім інформація передається до модуля керування інтерфейсом користувача.

З метою отримання якісно прогнозу детально розглянуто методи машинного навчання, визначено їх особливості. Встановлено, що ефективність класифікації можна покращити за допомогою комбінації класифікаторів і методів, таких як дерева рішень, метод опорних векторів та багатошаровий перцептрон. Саме така комбінація алгоритмів дозволить спрогнозувати оптимальний туристичний маршрут. Були розглянуті як технічні, так і практичні аспекти, включаючи розрідженість даних, масштабованість, прозорість, точність системи, зручність використання та прийнятність користувачами, що дозволило оцінити ефективність запропонованої методики.

Ключові слова: інформаційна технологія, туризм, машинне навчання, рекомендаційні системи, дерево рішень, метод опорних векторів, багатошаровий перцептрон.

Trehubchak I.M., Ilin O.Yu., Koba A.B.

State University of Telecommunications, Kyiv

INFORMATION TECHNOLOGY FOR CHOOSING ROUTES USING MACHINE LEARNING METHODS

Abstract. Information technologies penetrate into all spheres of human life and the activity of companies. The sphere of tourism did not escape their influence. Tourism is a leisure activity that involves complex decision-making processes, which is why the development of recommendation systems that will help facilitate these processes is relevant. It is for this purpose that the method of selecting tourist routes using machine learning methods was created in the work. In the course of the research, the field of tourism was analyzed, the peculiarities of recommendation systems were established. The levels of the proposed methodology are described, indicating the features of the backend and frontend. A characteristic feature of the proposed methodology is the use of machine learning classifiers, such as decision trees, support vector method, and multilayer perceptron, in the model of optimal selection. It is this combination of methods that will allow you to choose the best route from the proposed ones. They are used to distinguish specific destinations within each data set.

© Трегубчак І.М., Ільїн О.Ю., Коба А.Б.

2022

To make the complex model usable and interpret its results for the tourist, the decision tree models are transformed into decision rules, and then the information is passed to the user interface control module.

In order to obtain a high-quality forecast, the methods of machine learning were considered in detail, and their features were determined. It has been found that the classification performance can be improved by using a combination of classifiers and methods such as decision trees, support vector method and multilayer perceptron. This combination of algorithms will allow to predict the optimal tourist route. Both technical and practical aspects, including data sparsity, scalability, transparency, system accuracy, usability, and user acceptability, were considered to evaluate the effectiveness of the proposed methodology.

Keywords: *information technology, tourism, machine learning, recommender systems, decision tree, support vector method, multilayer perceptron.*

1. Постановка проблеми.

Останнім часом туризм отримав значну користь від інформаційних технологій, особливо від Інтернет-технологій та їх додатків. Інструменти підтримки прийняття рішень, також відомі як системи рекомендацій, були розроблені для вирішення цих проблем [5]. У сфері туризму вони називаються туристичними рекомендаційними системами. Туристи та постачальники туристичних послуг можуть шукати, вибирати, порівнювати та приймати рішення майже миттєво та ефективніше, ніж будь-коли.

2. Аналіз останніх досліджень і публікацій.

Через величезну кількість різномірної інформації, доступної в Інтернеті та через інші джерела інформації, туристичні рекомендаційні системи можуть діяти як інформаційні фільтри. Підбір відповідних туристичних послуг відповідно до вподобань користувачів є одним із найскладніших завдань, з якими стикається турист, плануючи поїздку в незнайоме місто. Незважаючи на те, що пошукові системи надають списки туристичних послуг, туристи все одно переповнені інформацією про пропозицію. Туристичні рекомендаційні системи можна широко використовувати як засіб зменшення інформаційного перевантаження для людей.

3. Мета і задачі дослідження.

Щоб надавати туристам практичну допомогу, туристичні рекомендаційні системи мають стати «розумними» щодо певних технічних аспектів, таких як масштабованість, прозорість, точність рекомендацій і методи перевірки; а також мати певні практичні аспекти, такі як сприйняття користувачами та зручність використання – усі вони повинні бути взяті до уваги при проектуванні системи. Крім того, ефективна туристична рекомендаційна система повинна досягти балансу між практичними та технічними аспектами.

Саме тому актуальним є покращення процесу вибору туристичних маршрутів на основі рекомендаційних систем з використанням методів машинного навчання. Покращення ефективності запропонованої методики забезпечується використанням алгоритмів класифікації таких як дерево рішень, метод опорних векторів та багатошаровий перцептрон.

Запропонована методика дозволить обрати найкращий маршрут для користувача, який буде відповідати визначеним критеріям.

4. Результати дослідження.

Запропонована методика може бути використана для допомоги туристам при їх підготовці до відвідування незнайомого міста. Вона призначена для використання туристами та турагентами та складається з декількох рівнів. На першому рівні система виконує розрахунок моделей оптимального вибору напрямків, щоб рекомендувати напрямки туристам. Через модуль керування даними необроблені дані, такі як записи оглядів, подаються в

систему. Цей модуль відповідає за інтеграцію, очищення, перетворення, зберігання та підтримку даних опитування. Обслуговування системи просто вимагає введення нових даних у систему рекомендацій у модулі керування даними на цьому рівні. При кожній інтеграції нових даних в існуючий набір даних створюються нові моделі, які передаватимуться на веб-сервер у верхньому рівні. У модулі керування інтерфейсом користувача можна додавати, редагувати, видаляти або змінювати моделі.

Другий рівень містить в собі клієнтський, оскільки в ньому є інтерфейс користувача, де турист може взаємодіяти з системою через різні платформи, такі як мобільний, робочий стіл або веб-браузер. У модулі керування інтерфейсом користувача правила прийняття рішень перетворюються у формати XML і JSON для створення нового інтерфейсу користувача. Крім того, система може підключатися до Google API для отримання відповідної інформації, яка стосується карт і маршрутів, щоб система могла відображати результати в інтерфейсі. Взаємодія користувачів з системою відбувається через інтерфейс користувача. Щоб отримати рекомендований пункт призначення, турист повинен надати ряд вхідних даних, наприклад мету поїздки та дохід користувача, а також інші характеристики в систему, вибравши відповіді зі списків. Рекомендовані результати включатимуть назву пункту призначення та маршрут подорожі, який буде отримано за допомогою інформації, отриманої з місця розташування користувача та вибраного пункту призначення. У цьому шарі зберігається географічна, просторова та маршрутна інформація. Система підключається до кількох API Google, таких як GMap і GLargeMap, щоб мати можливість завантажувати карти та керувати ними.

Характерною особливістю запропонованої методики є використання в моделі оптимального вибору класифікаторів машинного навчання, таких як дерева рішень, метод опорних векторів та багатошаровий перцептрон. Саме така комбінація методів дозволить обрати найкращий маршрут із запропонованих. Вони використовуються для розрізнення конкретних пунктів призначення в кожному наборі даних. Щоб зробити комплексну модель придатною для використання та інтерпретувати її результати для туриста, моделі дерева рішень перетворюються на правила прийняття рішень, а потім інформація передається до модуля керування інтерфейсом користувача.

Дерева рішень – це універсальні алгоритми машинного навчання, які можуть виконувати завдання класифікації і регресії з кількома виходами [1]. Це потужні алгоритми, які здатні обробляти складні набори даних.

Дерева рішень є також фундаментальними компонентами випадкових лісів, одних із найпотужніших алгоритмів машинного навчання. Однією з ключових особливостей дерев рішень є те, що вони не вимагають підготовки даних. Фактично вони взагалі не вимагають масштабування або центрування елементів. Дерева рішень також можуть оцінювати можливість того, що екземпляр належить певному класу k . Спочатку алгоритм переглядає дерево, щоб знайти ключовий вузол для цього екземпляра, а потім повертає співвідношення навчальних екземплярів класу k у цьому вузлі.

Дерева рішень мають багато переваг:

- вони прості для розуміння та інтерпретації,
- прості у використанні,
- універсальні;
- потужні.

Проте вони також мають кілька обмежень:

- деревам рішень подобаються ортогональні межі рішень, які роблять їх чутливими до ротації навчальних наборів.
- дерева рішень схильні до перенавчання під час роботи із завданнями регресії.

Метод опорних векторів

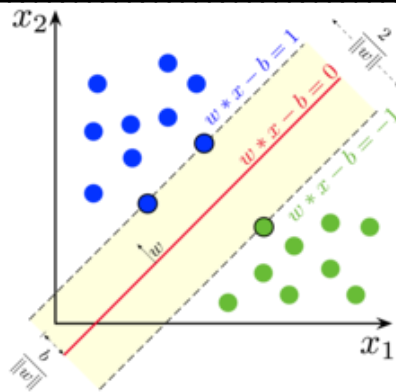


Рис. 1. Максимально розділова гіперплощина та межі [2]

Метод опорних векторів зазвичай використовується для вирішення проблем класифікації та регресії. Цей метод успішно застосовувався в багатьох областях для вирішення завдань класифікації, таких як розпізнавання цифрових символів рукописного тексту, виявлення обличчя тощо. Цей підхід проектує вхідні дані у простори з більшою вимірністю, щоб можна було розділити нелінійні дані. Метою є оптимізація гіперплощини, яку можна розділити на два вказаних класи об'єктів квадратами та колами, максимізуючи відстань кожної точки до гіперплощини, як показано на рис. 1. Метод опорних векторів складається з двох основних фаз. По-перше, функція ядра використовується для відображення даних у вищий вимір (тобто лінійна, поліноміальна, функція зміщення радіуса). У цей момент гіперплощина може бути використана для розділення двох класів. Для набору даних, який неможливо ідеально розділити лінійно, мета процесу полягає в тому, щоб знайти набір ваг, які визначають дві гіперплощини, як визначено нижче:

$$\begin{aligned} \bar{w} \cdot \bar{x} + b &\geq +1 \\ \bar{w} \cdot \bar{x} + b &\leq -1 \end{aligned} \quad (1)$$

У випадку нелінійно розділних даних метод опорних векторів може обробляти нероздільні точки, вводючи слабкі змінні, як показано нижче:

$$y_i (w^T x_i + b) \geq 1 - \xi_i \quad (2)$$

У цьому дослідженні гауссове ядро радіальної базисної функції було обрано як найбільш прийнятну функцію ядра, оскільки наш набір даних складається з невеликої кількості ознак, а радіальна базисна функція використовує менше гіперпараметрів, ніж поліноміальне ядро. Для цього дослідження було обрано RBF Гауса, як визначено в рівнянні:

$$f(x_i) = \exp\left(-\frac{1}{(2\sigma^2)} \|x_i - x_j\|^2\right) \quad (3)$$

Щодо переваг методу опорних векторів, то цей класифікатор здатний знаходити глобальний мінімум, а його проста геометрична інтерпретація створює необхідну базу для майбутніх досліджень. Найбільш вигідною характеристикою нелінійного класифікатора методу опорних векторів є опуклість. Однак метод опорних векторів також має кілька недоліків: він дуже чутливий до параметрів ядра та вибору ядра; отже, вибір параметра, який трохи виходить за межі, може призвести до низької ефективності класифікації.

Налаштування цих параметрів зазвичай необхідне для хорошої продуктивності. Наприклад, вибір параметра вартості є критичним. Використання більшої вартості може призвести до перепідгонки моделі. Крім того, розробка моделі за допомогою методу опорних векторів вимагає трудомісткого підходу проб і помилок і займає досить багато часу, особливо для великого обсягу даних. На рис.2 представлено алгоритм, який використовувався для оптимізації гіперпараметрів у методі опорних векторів у цьому дослідженні.

Багат шаровий персептрон

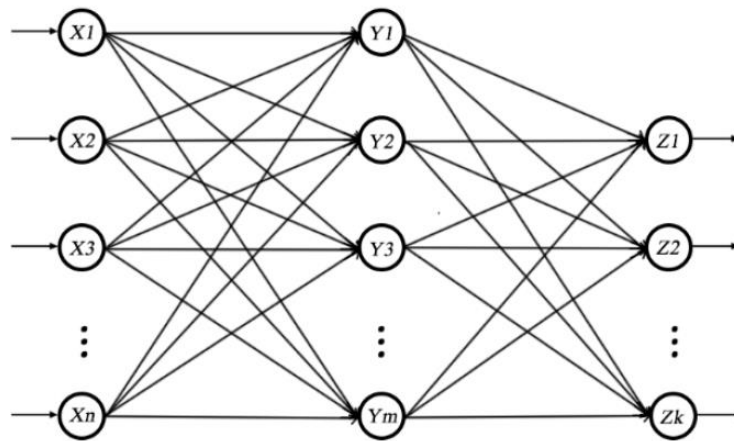


Рис.3. Архітектура багатошарового перцептрона з одним прихованим шаром

Багатошаровий перцептрон вважається прямою мережею, універсальним апроксиматором. Це найбільш поширений метод у галузі штучних нейронних мереж для вирішення завдань класифікації. Нейронну мережу можна навчити передбачати змінну класу. Існує багато типів штучних нейронних мереж, які використовуються для класифікації, включаючи багатошаровий перцептрон, радіальну базову функцію та ймовірнісні нейронні мережі. У цьому дослідженні як тип мережі було обрано саме багатошаровий перцептрон; її архітектура складається з одного або кількох прихованих шарів між вхідними та вихідними вузлами, і кожен із вузлів у мережі підключений і має певну вагу. Рис. 3 ілюструє загальну мережеву архітектуру перцептрону. Багатошаровий перцептрон відображає дані з простору ознак у вихідний простір класифікації, і прогноз можна вибрати як вектор кодування, який є найближчим до виходу (тобто вихід, який відображає найвище значення, є класом-переможцем).

Були розглянуті як технічні, так і практичні аспекти, включаючи розрідженість даних, масштабованість, прозорість, точність системи, зручність використання та прийнятність користувачами.

Запропонована система базується на трирівневій архітектурі веб-моделі, більш відомої як клієнт-серверна архітектура. Архітектура, яка складається з трьох рівнів, складається з рівня презентації, додатків і даних. Презентаційний рівень – це інтерфейс користувача, реалізований за допомогою технології веб-браузера, за допомогою якого він отримує вхідні дані, такі як демографічні дані, характеристики користувачів і вимоги користувачів від туристів, і відображає результати користувачам. Другий шар – це прикладний рівень, який виконує роль середнього шару. Він відповідає за оптимізацію та логічне прийняття рішень, а також за оцінку даних та інші розрахунки. Рівень даних приймає та зберігає всю інформацію з верхніх рівнів. Інформація та відповідні дані, як-от географічні дані та інформація про подорожі користувача, зберігаються на різних рівнях за допомогою форматів файлів eXtensible Markup Language (XML) і JavaScript Object Notation (JSON). У цьому дослідженні запропоновано інтерфейс користувача для запропонованої рекомендаційної системи, який має адаптивні, чутливі та інтерактивні можливості.

Адаптація для інтерфейсу користувача повинна включати деякі фактори, такі як продуктивність користувача, цілі користувача, когнітивне навантаження, обізнаність користувача про ситуацію, знання користувача, профілі груп, змінні ситуації та змінні завдання. Дерево рішень можна використовувати як алгоритм адаптації та як один із методів адаптації інтерфейсу. В інтерфейсі користувача швидкість реагування стосується змін розміру вікна браузера та способу розташування вмісту.

Інтерактивність є одним із найбільш багатообіцяючих аспектів, які слід розглянути, щоб використати весь потенціал рекомендаційної системи. Розробка та впровадження справжнього інтерактивного веб-сайту потребує великої роботи, яка передбачає спільне ставлення користувачів, чіткий процес і стандарти для управління вмістом, а також дослідження дизайну.

У цьому дослідженні було поставлено завдання збільшити інтерактивність між користувачем і системою, щоб відображати корисну інформацію (наприклад, місце призначення) користувачам за допомогою інтерактивних карт. Крім того, передові веб-технології, такі як JQuery, CSS і HTML5, можуть бути використані для покращення взаємодії з користувачем і підвищення відгуку та інтерактивності системи.

Запропонована рекомендаційна система призначена для використання туристами та турагентами та складається з онлайн- та офлайн-фаз. У офлайн-фазі система виконує розрахунок моделей оптимального вибору напрямків, щоб рекомендувати напрямки туристам, заощаджуючи туристам додаткові витрати на використання обладнання та час під час процесу пошуку інформації. Необроблені дані, наприклад записи обстежень, подаються в систему через модуль керування даними. Цей модуль відповідає за інтеграцію, очищення, перетворення, зберігання та підтримку даних опитування. Обслуговування системи просто вимагає введення нових даних у систему рекомендацій у модулі керування даними на цьому рівні. Наприклад, щороку, коли отримуються нові дані опитування, їх можна інтегрувати у існуючий набір даних, і відповідно створюватимуться нові моделі, які передаватимуться на веб-сервер у верхньому рівні. У модулі керування інтерфейсом користувача можна додавати, редагувати, видаляти або змінювати моделі.

У модулі «Керування моделлю» встановлено класифікатори дерево рішень та інші класифікатори машинного навчання, включаючи три добре відомі алгоритми класифікації: дерево рішень, метод опорних векторів, багат шаровий перцептрон та інші моделі ансамблевого навчання. Вони використовуються для розрізнення конкретних пунктів призначення в кожному наборі даних. Щоб зробити комплексну модель придатною для використання та інтерпретувати її результати для туриста, моделі дерева рішень перетворюються на правила прийняття рішень, а потім інформація передається до модуля керування інтерфейсом користувача (рис. 4).

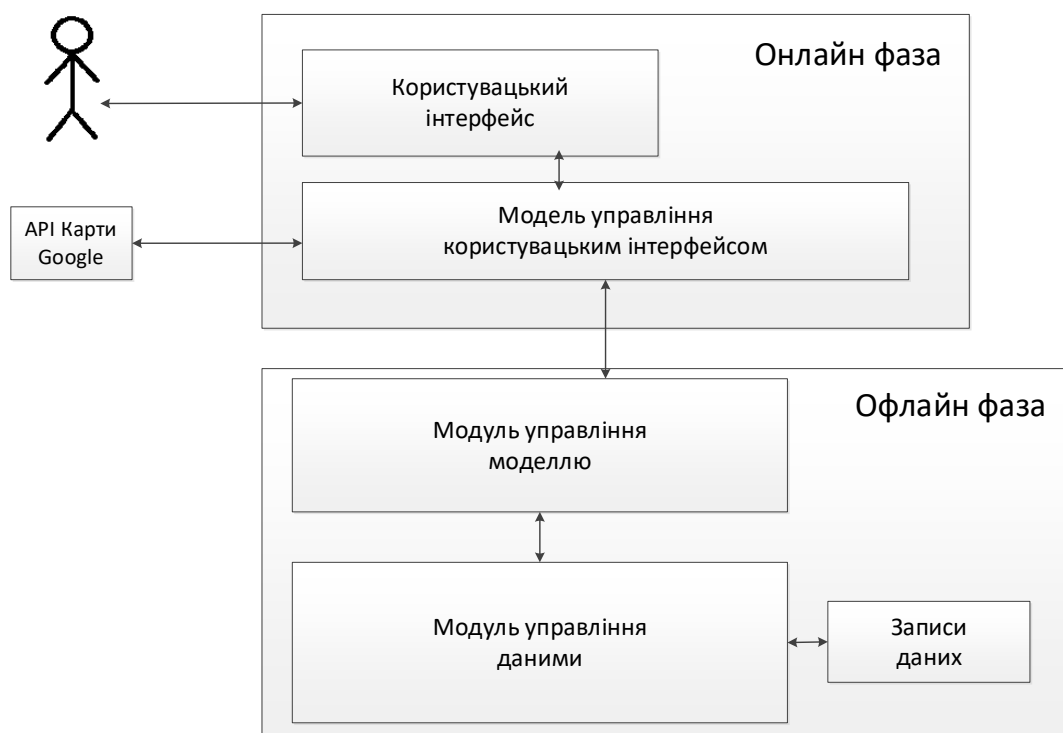


Рис.4. Практична система рекомендацій для туристів

На онлайн-фазі верхній рівень можна вважати клієнтським, оскільки він містить інтерфейс користувача, де турист може взаємодіяти з системою через різні платформи, такі як мобільний, робочий стіл або веб-браузер. У модулі керування інтерфейсом користувача правила прийняття рішень перетворюються у формати XML і JSON для створення нового

інтерфейсу користувача. Крім того, система може підключатися до Google API для отримання відповідної інформації, яка стосується карт і маршрутів, щоб система могла відображати результати в інтерфейсі. Туристи можуть взаємодіяти з системою через інтерфейс користувача. Щоб отримати рекомендований пункт призначення, турист повинен надати ряд вхідних даних, наприклад стиль поїздки та межі вартостей, а також інші в систему, вибравши відповіді зі списків. Згодом рекомендовані результати включатимуть назву пункту призначення та маршрут подорожі, який буде отримано за допомогою наданої інформації, отриманої з місця розташування користувача та вибраного пункту призначення. У цьому шарі зберігається географічна, просторова та маршрутна інформація. Система підключається до кількох API Google, таких як GMap і GLargeMap, щоб мати можливість завантажувати карти та керувати ними.

Після проведення відповідних досліджень та обчислень було розроблено інформаційну технологію рекомендації туристичних маршрутів, яка має наступну архітектуру (рис.4).

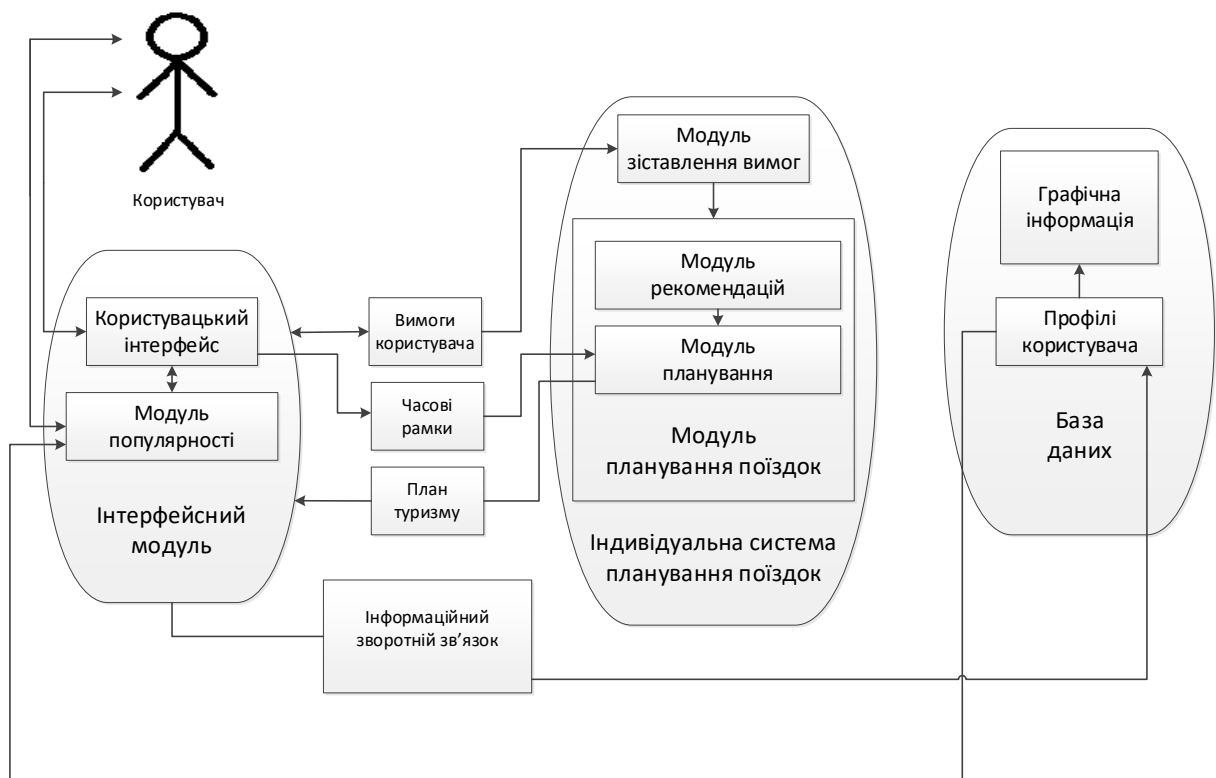


Рис.5. Загальна структура інформаційної технології

Як показано на рис. 5, модулі, на яких було зосереджено це дослідження, це модуль бази даних і модуль індивідуальної системи планування подорожей, а також алгоритм обґрунтування розкладу, який використовується для створення персоналізованого розкладу подорожей із кінцевого набору туристичних послуг, що включають місця визначних пам'яток, заклади харчування та ресторани, варіанти розміщення, розташування готелів, вимоги користувачів тощо. Алгоритм включає кілька кроків для пошуку місця подорожі чи пункту призначення та розрахунки, пов'язані з транспортуванням і часом перебування. Механізм зворотного зв'язку – це метод, який використовується для ранжування цікавих місць (готелі, ресторани, житло), який є сукупним значенням рейтингів популярності користувачів.

Модуль зіставлення вимог щодо подорожі відповідає введенням користувачам (наприклад, необхідним пам'яткам, готелям, ресторанам із бази даних). Потім рекомендований модуль виконав дії:

1. Пошук місця подорожі або пункту призначення.
2. Розрахунок транспортування та часу проживання.

3. Додавання обраного місця подорожі в часові рамки.
Інтерфейс рекомендаційної системи представлений на рис. 6-7.

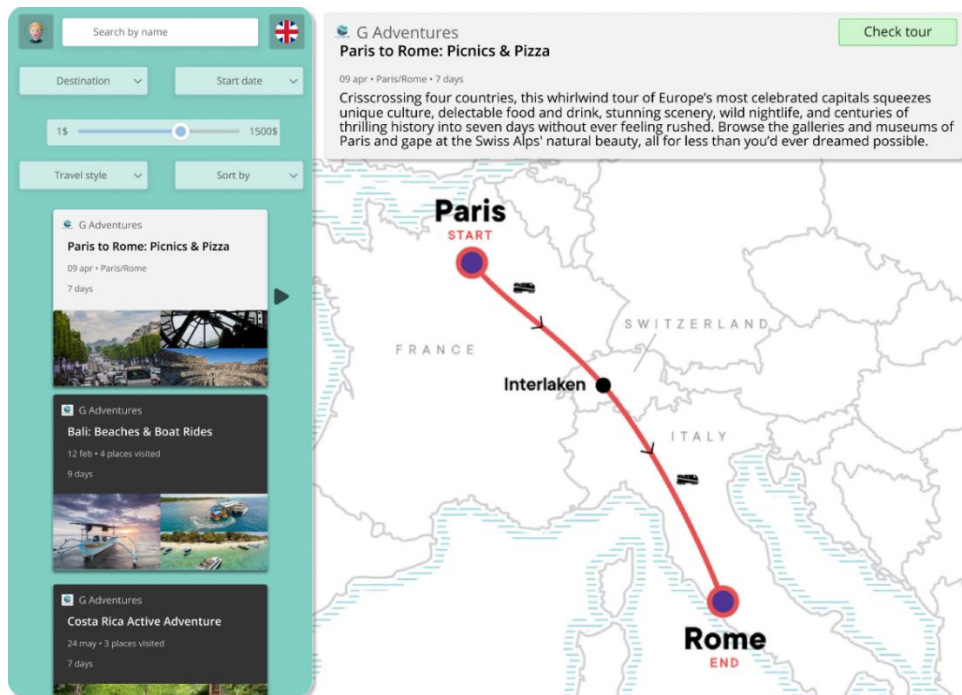


Рис.6. Виведення деталей обраного варіанту

В додатку є можливість вибору мови інтерфейсу, що полегшує користування додатком жителю будь-якої країни.

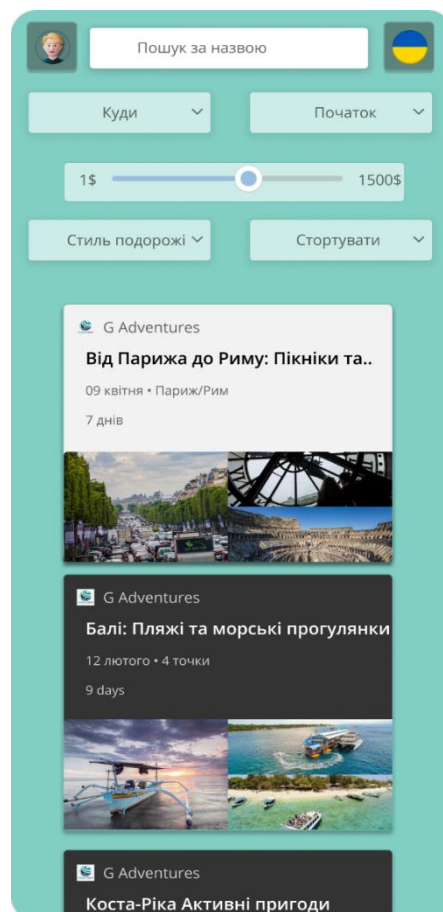


Рис.7. Мобільна версія додатку

Отже запропонована інформаційна технологія допомагає оптимізувати роботу туроператора та зменшити час при виборі туристичного маршруту самим туристом.

5. Висновки і перспективи подальших досліджень.

Відповідно до показників оцінки можна побачити, що різні алгоритми класифікації краще класифікують різні набори даних вибору призначення. Ефективність класифікації можна покращити за допомогою комбінації цих класифікаторів і методів сукупності класифікаторів. В результаті застосування запропонованої методики було отримано точність прогнозування 80,73%.

Ґрунтуючись на введенні користувачами даних запропонована методика може:

- 1) рекомендувати результати, що ґрунтуються на оцінці інтересу користувача;
- 2) рекомендувати об'єкти інтересу, туристичні послуги або маршрути;
- 3) ранжувати запропоновані напрямки;
- 4) запропонувати цілісний план поїздки.

Сферою подальших досліджень є застосування рекомендаційних систем та запропонованої методики до інших напрямків людської діяльності .

Список використаних джерел

1. Christopher M Bishop. Pattern recognition. Machine Learning, 2006, 128 p.
2. Ethem Alpaydin. Introduction To Machine Learning, 2009, 584 p.
3. Garcia I., Sebastia L., Onaindia E. On the design of individual and group recommender systems for tourism. Expert Systems with Applications 38 - 2011. 7683–7692. <https://doi.org/10.1016/j.eswa.2010.12.143>
4. Gavalas D., Kenteris M., Konstantopoulos C., Pantziou G. Web application for recommending personalised mobile tourist routes. IET Software 6 - 2012. 313–322. <https://doi.org/10.1049/iet-sen.2011.0156>
5. Isinkaye F.O., Folajimi Y.O., Ojokoh B.A. Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal 16 -2015. 261–273. <https://doi.org/10.1016/j.eij.2015.06.005>
6. Kantamneni A., Brown L.E., Parker G., Weaver W.W. Survey of multi-agent systems for microgrid control. Engineering Applications of Artificial Intelligence 45 - 2015. 192–203. <https://doi.org/10.1016/j.engappai.2015.07.005>
7. Saleh E., Błaszczyński J., Moreno A., Valls A., Romero-Aroca P., de la Riva-Fernández S., Słowiński R. Learning ensemble classifiers for diabetic retinopathy assessment. Artificial Intelligence in Medicine - 2017. <https://doi.org/10.1016/j.artmed.2017.09.006>
8. Tom M. Mitchell. Machine Learning [<http://www.cs.cmu.edu/~tom/mlbook.html>]
9. Yeh, D.-Y., Cheng, C.-H. Recommendation system for popular tourist attractions in Taiwan using Delphi panel and repertory grid techniques. Tourism Management 46 - 2015. 164–176. <https://doi.org/10.1016/j.tourman.2014.07.002>

References:

1. Christopher M Bishop. Pattern recognition. Machine Learning, 2006, 128 p.
2. Ethem Alpaydin. Introduction To Machine Learning, 2009, 584 p.
3. Garcia I., Sebastia L., Onaindia E. On the design of individual and group recommender systems for tourism. Expert Systems with Applications 38 - 2011. 7683–7692. <https://doi.org/10.1016/j.eswa.2010.12.143>
4. Gavalas D., Kenteris M., Konstantopoulos C., Pantziou G. Web application for recommending personalised mobile tourist routes. IET Software 6 - 2012. 313–322. <https://doi.org/10.1049/iet-sen.2011.0156>

5. Isinkaye F.O., Folajimi Y.O., Ojokoh B.A. Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal 16 -2015. 261–273. <https://doi.org/10.1016/j.eij.2015.06.005>
6. Kantamneni A., Brown L.E., Parker G., Weaver W.W. Survey of multi-agent systems for microgrid control. Engineering Applications of Artificial Intelligence 45 - 2015. 192–203. <https://doi.org/10.1016/j.engappai.2015.07.005>
7. Saleh E., Błaszczński J., Moreno A., Valls A., Romero-Aroca P., de la Riva-Fernández S., Słowiński R. Learning ensemble classifiers for diabetic retinopathy assessment. Artificial Intelligence in Medicine - 2017. <https://doi.org/10.1016/j.artmed.2017.09.006>
8. Tom M. Mitchell. Machine Learning [<http://www.cs.cmu.edu/~tom/mlbook.html>]
9. Yeh, D.-Y., Cheng, C.-H. Recommendation system for popular tourist attractions in Taiwan using Delphi panel and repertory grid techniques. Tourism Management 46 - 2015. 164–176. <https://doi.org/10.1016/j.tourman.2014.07.002>