

MINISTERUL EDUCAȚIEI ȘI CERCETĂRII AL REPUBLICII MOLDOVA
Universitatea Tehnică a Moldovei
Facultatea Calculatoare, Informatică și Microelectronică
Departamentul Ingineria Software și Automatică

Admis la susținere
Șef departament:
FIODOROV Ion dr., conf.univ.

”___” _____ 2024

ANALIZA ȘI IMPLEMENTAREA TEHNOLOGICĂ A SISTEMELOR DE
RECOMANDARE ÎN STREAMING ȘI MAGAZINE ONLINE

Proiect de master

Student: _____ **Vorobiov Daniel, TI-221M**

Coordonator: _____ **Cojocaru Svetlana, asist.univ.**

Consultant: _____ **Cojocaru Svetlana, asist.univ.**

Chișinău, 2024

REZUMAT

Lucrarea respectivă explorează domeniul recomandărilor pentru o experiență cât mai organică online al unui utilizator. În contextul acestui studiu se va cerceta conceptul de sisteme de recomandare de conținut digital din punct de vedere tehnologic. Se va aprecia importanța existenței recomandărilor atât pentru un utilizator, cât și impactul economic a unui astfel de sistem. La fel, se va cerceta necesitatea personalizării recomandărilor, și beneficiile cât a utilizatorilor atât și a producătorilor de conținut digital și de vânzători online.

Considerând că în baza modelelor ce generează aceste recomandări se află un set de algoritmi clasici combinați cu elemente de Machine Learning, lucrarea respectivă va analiza și defini cei mai des întâlniți algoritmi în contextul sistemelor de recomandare. La fel se vor studia și instrumentele utilizate pentru a crea un astfel de sistem, în fundamentele cărora se află limbajul de programare Python. Librăriile din acest limbaj sunt cele mai rapide și puternice în crearea modelelor matematice de diferite tipuri. Cu ajutorul acestor librării se va efectua și o analiză exploratorie a datelor care va oferi mai mult context și detalii în legătura cu sistemele de recomandare ce urmează a fi realizate.

Lucrarea va analiza 2 seturi de date: unul cu feedback implicit și cu altul feedback explicit, pentru a putea forma o opinie despre provocările fiecăruia. În dependență de set, se vor analiza modelele care ar putea fi aplicate în crearea unui sistem de recomandare. Cu ajutorul acestor modele urmează să fie create 2 sisteme de recomandare. Primul va fi dedicat recomandărilor de filme și conținut digital utilizând date cu feedback explicit. Al doilea sistem va fi responsabil să genereze recomandări de produse utilizatorilor unui magazin online în baza unui set de date cu feedback implicit care va analiza cât activitatea utilizatorului atât și a al altor utilizatori cu interese similare.

ABSTRACT

This study explores the domain of recommendation system for a user's most organic online experience. In the context of this study, the concept of digital content recommendation systems will be investigated from a technological point of view. The importance of the existence of recommendations for both a user and the economic impact of such a system will be appreciated. Likewise, the need for personalization of recommendations will be researched, and the benefits to users, digital content producers and online sellers.

Considering that the models that generate these recommendations are based on a set of classic algorithms combined with Machine Learning elements, the respective paper will analyze and define the most common algorithms in the context of recommendation systems. Likewise, the tools used to create such a system will be studied, the foundations of which are the Python programming language. Libraries in this language are the fastest and most powerful in creating mathematical models of various types. With the help of these libraries, an exploratory analysis of the data will also be carried out, which will provide more context and details in connection with the recommendation systems to be implemented.

The paper will analyze 2 sets of data: one with implicit feedback and another with explicit feedback, to be able to form an opinion about the challenges of each. Depending on the dataset, models that could be applied in creating a recommender system will be analyzed. With the help of these models, 2 recommendation systems are to be created. The first will be dedicated to movies and digital content recommendations using data with explicit feedback. The second system will be responsible for generating product recommendations to users of an online store based on an implicit feedback dataset that will analyze both the user's activity and that of other users with similar interests.

CUPRINS

INTRODUCERE	8
1 ANALIZA DOMENIULUI DE STUDIU	9
1.1 Platformele de streaming	9
1.2 Magazinele online	11
1.3 Scopuri și obiective.....	12
2 ALGORITMI	14
2.1 Algoritmi liniari	14
2.2. Algoritmi recursivi.....	14
2.3. Algoritmi de căutare	15
2.4 Algoritmi de sortare	17
2.5. Algoritmi de învățare automată	18
2.6. Algoritmi de recomandare	19
3 SISTEM DE RECOMANDARE PENTRU O PLATFORMA DE STREAMING.....	21
3.1 Instrumente și tehnologii utilizate.....	21
3.2 Pregătirea modelului	23
3.3 Îmbunătățirea modelului	28
3.4 Prezicerile modelului	30
4 SISTEM DE RECOMANDARE PENTRU UN MAGAZIN ONLINE	33
4.1 Pregătirea modelului	33
4.2 Modelul LightFM.....	35
4.3 Modelul regresiei logistice.....	38
CONCLUZII	42
BIBLIOGRAFIE.....	43

INTRODUCERE

Această lucrare reprezintă o investigație detaliată și cuprinzătoare în domeniul algoritmilor de recomandare, concentrându-se în mod special pe eficiența lor în contextul platformelor de streaming de conținut digital și al site-urilor de magazine online. Aceste două domenii reprezintă nu doar o majoritate a conținutului web, ci și două aspecte cruciale ale experienței utilizatorului și ale economiei digitale.

Algoritmii de recomandare, coloana vertebrală a experienței personalizate online, sunt motorul care direcționează ceea ce vedem, ascultăm sau cumpărăm în mediul digital. Acești algoritmi folosesc o combinație complexă de tehnici și modele matematice pentru a analiza istoricul utilizatorilor, preferințele lor, comportamentul de navigare și alte date relevante. Cu aceste informații, ei creează recomandări personalizate care să prezinte utilizatorilor conținut sau produse care se potrivesc cel mai bine gusturilor și nevoilor lor.

Pentru a atinge obiectivele acestei investigații, se va baza pe o serie de instrumente tehnologice avansate. Machine Learning și Deep Learning, două ramuri importante ale inteligenței artificiale, vor fi folosite pentru a analiza și a compara performanța diferitelor algoritmi de recomandare. Aceste abordări moderne se bazează pe capacitatea de a învăța din datele de antrenament și de a face previziuni precise și relevante.

Cu toate acestea, nu se va neglija nici fundamentele matematice care stau la baza acestor tehnologii. Statistica, algebra liniară, teoria probabilităților și altele asemenea reprezintă bazele solide pe care se construiesc acești algoritmi. Prin înțelegerea lor profundă, voi putea evalua în mod critic performanța algoritmilor de recomandare și voi identifica posibilele limitări sau bias-uri.

Această lucrare se angajează să ofere o analiză completă a eficienței algoritmilor de recomandare în contextul platformelor de streaming de conținut digital și al site-urilor de magazine online. Prin explorarea acestor aspecte complexe și prin utilizarea unui set diversificat de instrumente tehnologice și matematice, îmi propun să contribui la dezvoltarea unei înțelegeri mai profunde și a unor practici mai bune în acest domeniu esențial al mediului digital contemporan.

BIBLIOGRAFIE

- [1] „Binary Search in Java - GeeksforGeeks”. Data accesării: 7 ianuarie 2024. [Online]. Disponibil la: <https://www.geeksforgeeks.org/binary-search-in-java/>
- [2] K. Gupta, „What is the Difference Between BFS and DFS Algorithms”, Developers, Designers & Freelancers – FreelancingGig. Data accesării: 7 ianuarie 2024. [Online]. Disponibil la: <https://www.freelancinggig.com/blog/2019/02/06/what-is-the-difference-between-bfs-and-dfs-algorithms/>
- [3] „Depth First Search or DFS for a Graph - GeeksforGeeks”. Data accesării: 7 ianuarie 2024. [Online]. Disponibil la: <https://www.geeksforgeeks.org/depth-first-search-or-dfs-for-a-graph/>
- [4] „Selection Sort (With Code in Python/C++/Java/C)”. Data accesării: 7 ianuarie 2024. [Online]. Disponibil la: <https://www.programiz.com/dsa/selection-sort>
- [5] „What is Supervised Learning? | IBM”. Data accesării: 7 ianuarie 2024. [Online]. Disponibil la: <https://www.ibm.com/topics/supervised-learning>
- [6] „What is a Neural Network? - Artificial Neural Network Explained - AWS”. Data accesării: 7 ianuarie 2024. [Online]. Disponibil la: <https://aws.amazon.com/what-is/neural-network/>
- [7] „Collaborative Filtering Vs Content-Based Filtering for Recommender Systems”. Data accesării: 7 ianuarie 2024. [Online]. Disponibil la: <https://analyticsindiamag.com/collaborative-filtering-vs-content-based-filtering-for-recommender-systems/>
- [8] „Introduction To Recommender Systems- 1: Content-Based Filtering And Collaborative Filtering | by Abhijit Roy | Towards Data Science”. Data accesării: 7 ianuarie 2024. [Online]. Disponibil la: <https://towardsdatascience.com/introduction-to-recommender-systems-1-971bd274f421>