



Algoritmy odporúčania filmov v Netflix*

Yehor Bohuslavskiy

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií
xbohuslavskiy@stuba.sk

17. október 2024

Abstrakt

Túto tému som si vybral, pretože je veľmi zaujímavé, ako technológie a umelá inteligencia (ktorá je v súčasnosti jednou z najrozvíjajúcejších sa tém) ovplyvňujú naše každodenné rozhodovanie o tom, čo sledujeme. Ja som zvedavý, že ako po zhliadnutí jedného tureckého filmu, ktorý som ani neohodnotil, sa mi v odporúčaní začali objavovať podobné filmy.

V systéme sa nachádza viac, ako 15 000 filmov a zobrazuje vám len tie, ktoré sa vám budú páčiť (neskôr dozvieme sa, ako to funguje). Fakt: ani jeden používateľ Netflix nebol by schopný si samostatne nájsť film alebo seriál, ktorý by sa mu páčil bez požívania algoritmu odporúčania filmov [1].

V tomto článku sa podrobne pozrieme na algoritmy spoločnosti Netflix. V súčasnosti je systém Netflix postavený na algoritme, ktorý využíva umelú inteligenciu a strojové učenie. Pozrieme sa aj na umiestnenie odporúčaní na obrazovke, kde je niekoľko typov návrhov. Prvým je „Odporúčané pre vás“, po ňom idú „Trendy“, na stránke sa nachádza aj záložka „New Releases“ [2].

Podrobne sa pozrieme aj na filtračné algoritmy, ktoré sa delia na dva hlavné typy. Akú rolu odohrávajú hodnotenia a správanie používateľov v rámci algoritmu „collaborative filtering“ a „filtrovanie na základe obsahu“.

Zaujímavým aspektom sú aj systémy hodnotenia. Patrí medzi ne „Personalizované hodnotenie videa (PVR)“, „Top-N Ranking“, „Rebríček obľúbených filmov“ a „Zoznam zaujímavého obsahu na neskoršie sledovanie“. Každé z týchto hodnotení je personalizované pre jednotlivých používateľov, čo je veľmi zaujímavé.

V rámci článku sa budeme venovať aj typom údajov, ktoré Netflix používa. Tie sa zhruba delia na dva typy: základné typy zhromažďovaných údajov a dodatočné informácie. Medzi základné patria „interakcie používateľa so stránkou“, ktoré pozostávajú z histórie sledovania a hodnotení. Druhým typom sú „korelačné údaje“, „informácie o obsahu knižnice Netflix“, ako napríklad žánre. Ďalšie uvidíme ako ovplyvňujú špecifickejšie typy informácií, ako sú napríklad denný čas, používané zariadenie alebo priemerná dĺžka sledovania.

*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2024/25, vedenie: Ing. Richard Marko, PhD

1 Úvod

Cieľom spoločnosti Netflix je používať odporúčacie algoritmy na maximalizáciu spokojnosti používateľov. Hlavným cieľom je prispôsobiť obsah každému jednotlivému používateľovi.

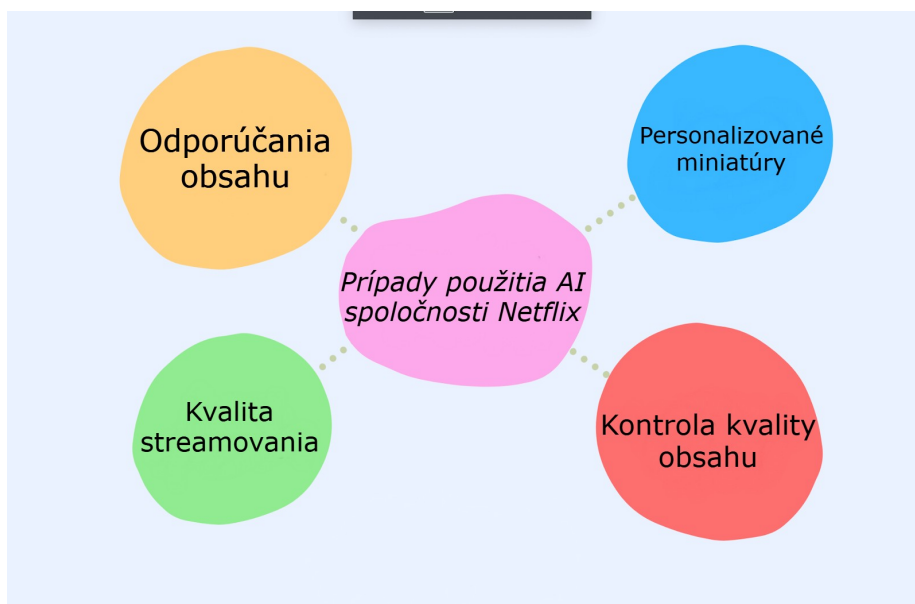
Netflix má obrovskú knižnicu obsahu a obmedzené rozhranie obrazovky, čo veľmi sťažuje odporúčanie relevantného obsahu, a každý používateľ má iný vkus, preferencie a zvyky pri sledovaní, takže personalizácia musí byť veľmi presná. [3] Čo robí Netflix na dosiahnutie tohto cieľa, sa dozviete po tomto článku.

2 Technológie a Umelá inteligencia v Netflix

Netflix využíva umelú inteligenciu a strojové učenie na zlepšenie používateľského zážitku. Medzi kľúčové aplikácie týchto technológií patrí

1. **Odporúčania obsahu:** na základe vašich zvykov a preferencií vám algoritmy spoločnosti Netflix môžu ponúknuť personalizované odporúčania. Analýza údajov zahŕňa aj žáner a obľúbenosť obsahu [4].
2. **Personalizované miniatúry:** Ak si zoberieme ako príklad film „Shadow and Bone“, miniatúry filmu sú prispôbené vašim záujmom a môžu používať obrázky súvisiace s inými populárnymi programami, skrátka čokoľvek, čo upúta vašu pozornosť. [4]
3. **Kvalita streamovania:** Na zabezpečenie plynulého prehrávania videa využíva Netflix údaje na predpovedanie preťaženia siete a ukladá obsah do medzipamäte na blízkych serveroch, takže si môžete vychutnať filmy bez oneskorenia. [4] [5]
4. **Kontrola kvality obsahu:** Pred odporúčaním nového obsahu používateľom prechádza niekoľkými fázami testovania vrátane zvuku a obrazu, aby boli všetci používatelia spokojní. [4] [6]

Táto technológia nielenže zvyšuje spokojnosť používateľov, ale podporuje aj dlhodobú angažovanosť.



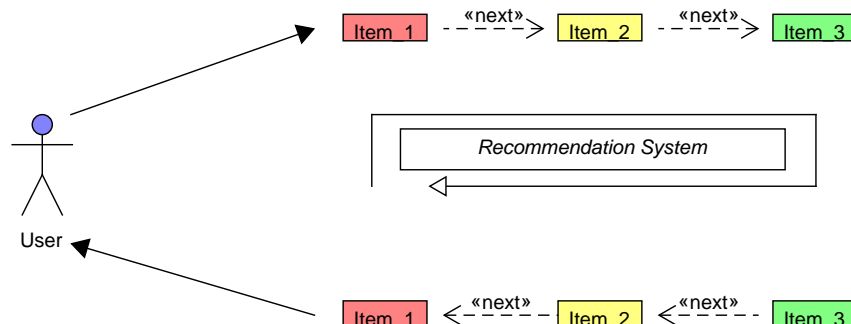
Obr. 1: Systém umelej inteligencie

3 Typy Odporúčaní na Netflix

1. **Odporúčania obsahu:** na základe vašich zvykov a preferencií vám algoritmy spoločnosti Netflix môžu ponúknuť personalizované odporúčania. Analýza údajov zahŕňa aj žáner a obľúbenosť obsahu [7].
2. **Kolaboratívne filtrovanie:** vytvorením skupín používateľov, ktorí majú podobný vkus, a na základe správania a preferencií používateľov táto metóda navrhuje obsah, ale má problém so studeným štartom. [7] [8]
3. **Filtrovanie na základe obsahu:** tento typ filtrovania zhromažďuje údaje, napríklad aký žáner filmu preferujete, akého režiséra alebo akých hercov uprednostňujete, a odporúča vám podobný obsah. [7]
4. **Hybridné modely:** Tento typ kombinuje kolaboratívne filtrovanie a filtrovanie založené na obsahu, pretože len tak sa dá vyriešiť problém studeného štartu. [7]
5. **Odporúčania založené na hlbokom učení:** Používajú sofistikované algoritmy umelej inteligencie, ktoré modelujú interakciu používateľa ako postupný rozhodovací proces. [7] [9]

4 Filtračné algoritmy

Netflix používa rôzne typy filtrovania, ako je kolaboratívne filtrovanie, filtrovanie na základe obsahu, hybridné modely alebo odporúčania založené na hlbokom



Obr. 2: Recommendation System

učení, ale v tomto článku sa budeme zaoberať len prvými dvoma, pretože sú najpoužívannejšie. Rozdiel medzi nimi spoznáte podľa názvu, ale poďme sa na ne pozrieť bližšie. [10]

Kolaboratívne filtrovanie je algoritmus, ktorý sleduje, aké sú vaše preferencie pri výbere filmov, a na základe týchto údajov vás zoskupí s ostatnými ľuďmi, ktorí majú podobný vkus. Názornejšie, ako funguje, nájdete na Obr.3, kde je graficky znázornená metóda kolaboratívneho filtrovania. Hoci sú na obrázku len 2 ľudia, v skutočnosti ich môže byť 100 alebo viac. Hlavným problémom tohto systému je problém studeného štartu, ktorý sa objavuje ako dôsledok nedostatku údajov o používateľovi alebo filmoch v systéme. Tento problém sa rieši použitím niekoľkých typov filtrovania, ale Netflix sa naďalej snaží tento problém vyriešiť. [11] Metóda kolaboratívneho filtrovania sa delí na dve ďalšie metódy, prvou je **Kolaboratívne filtrovanie založené na používateľovi** (v tabuľke označené ako UBCF), zameriava sa na vyhľadávanie ľudí s podobným vkusom ako vy. Ako presne? Najprv vypočíta podobnosť medzi všetkými používateľmi a v závislosti od výsledku začne odporúčať filmy, ktoré ste ešte nevideli, pričom sa páčia ľuďom s podobným vkusom. [12] Druhou čiastkovou metódou je **Kolaboratívne filtrovanie založené na položkách** (IBCF), líši sa tým, že nehľadá podobnosti medzi ľuďmi, ale medzi položkami, s ktorými ste komunikovali, a odporúča podobné. Ak sa vám napríklad páči film, v ktorom hral Robert Downey Jr, systém vám začne odporúčať filmy alebo seriály, v ktorých hral tento herec. Podrobnejšie, aký je medzi nimi rozdiel a aké aspekty zahŕňajú, si môžete pozrieť v nasledujúcej tabuľke. [12]

Filtrovanie na základe obsahu tento typ filtrovania je veľmi ľahko pochopiteľný, pretože podstatou algoritmu je odporučiť obsah, ktorý je podobný predchádzajúcim výberom používateľa (obrázok 3). Algoritmus venuje pozornosť charakteristikám, ako sú „názov“, „krajina“, „režisér“, „obsadenie“, „uveďené“ a „opis“ filmu/seriálu, aby odporučil filmy, ktoré zodpovedajú vášmu vkusu. Ak napríklad radi sledujete komédie, Netflix vám začne odporúčať filmy s Jackie Chanom alebo komédie s inými populárnymi hercami. Algoritmus tiež uvidí, že máte radi filmy s Jackie Chanom, a začne odporúčať filmy súvisiace s východnou kultúrou, a tak Netflix postupne zachytí vaše preferencie a bude sa

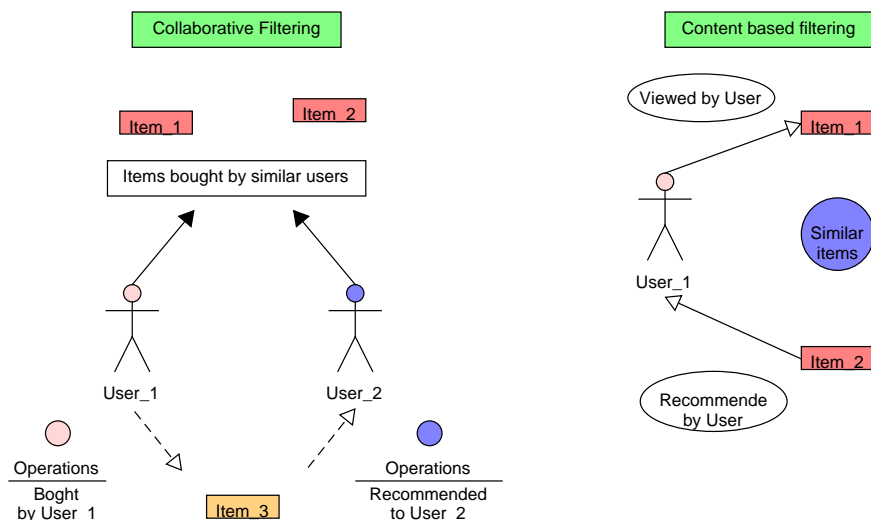
#	Aspekt	Kolaboratívne filtrovanie na základe používateľov (UBCF)	Kolaboratívne filtrovanie na základe položiek (IBCF)
1	Zameranie	Odporúčania sú založené na podobnosti medzi používateľmi.	Odporúčania sú založené na podobnosti medzi položkami (produkty, filmy, atď.).
2	Vstupné dáta	Matica hodnotení používateľov (riadky reprezentujú používateľov, stĺpce reprezentujú položky).	Matica hodnotení používateľov (riadky reprezentujú používateľov, stĺpce reprezentujú položky).
3	Cieľ odporúčania	Na odporúčanie sa používajú používatelia, ktorí sú podobní cieľovému používateľovi.	Odporúčajú sa položky podobné tým, ktoré cieľový používateľ hodnotil alebo s nimi interagoval.
4	Logika algoritmu	1. Vypočíta sa podobnosť medzi cieľovým používateľom a všetkými ostatnými používateľmi. 2. Identifikujú sa používatelia s podobnými preferenciami ako cieľový používateľ. 3. Odporúčajú sa položky, ktoré podobní používatelia hodnotili, ale cieľový používateľ ešte nie.	1. Vypočíta sa podobnosť medzi všetkými párami položiek. 2. Identifikujú sa položky podobné tým, ktoré cieľový používateľ hodnotil alebo s nimi interagoval. 3. Odporúčajú sa podobné položky cieľovému používateľovi.
5	Výpočet podobnosti	Typicky používa metriky ako kosínová podobnosť alebo Pearsonova korelácia medzi hodnoteniami používateľov.	Typicky používa metriky ako kosínová podobnosť alebo Jaccardova podobnosť medzi hodnoteniami položiek.
6	Riedkosť dát	UBCF môže byť menej účinný s riedkymi dátami, kde používatelia hodnotili len niekoľko položiek.	IBCF môže lepšie zvládať riedke dáta, keďže vzťahy medzi položkami sú zvyčajne hustejšie.
7	Škálovateľnosť	UBCF môže byť výpočtovo náročné, ak sa zvyšuje počet používateľov.	IBCF je často viac škálovateľný, najmä pri práci s veľkým množstvom položiek.
8	Príklady použitia	1. Odporúčania filmov na základe preferencií používateľov. 2. Odporúčania produktov na e-commerce platformách.	1. Odporúčanie podobných produktov k aktuálne prezeranému položke používateľa. 2. Odporúčanie súvisiacich článkov v spravodajskej aplikácii.

Obr. 3: UBCF/IBCF

stále zlepšovať. [13]

5 Systémy Hodnotenia

Kľúčovými zložkami odporúčacieho systému spoločnosti Netflix je niekoľko personalizovaných systémov hodnotenia, z ktorých každý je navrhnutý tak, aby zohľadňoval rôzne aspekty zvykov používateľa pri sledovaní. Poďme sa do týchto systémov ponoriť a preskúmať, ako fungujú.



Obr. 4: Type of Filtering

1. **Personalizovaný rebríček videí (PVR)** je to jeden z najsofistikovanejších nástrojov, ktoré používa spoločnosť Netflix, pretože dokáže nielen pochopiť minulé správanie, ale aj predpovedať budúce interakcie. [14] Pozrime sa bližšie na to, ako funguje, a začnime tým, že je založený na systéme hlbokého učenia a neustále sa prispôsobuje používateľovi. [15] Hlavnou myšlienkou je využitie historických údajov, napríklad zvykov používateľa, na identifikáciu vašich zvykov a uprednostnenie obsahu, ktorý vás bude zaujímať. Celkovo poskytuje celkové relatívne poradie položiek v rade s využitím personalizovaných údajov používateľa. [16]
2. **Systém Top-N Ranking** je založený na kombinácii kolaboratívneho a obsahového filtrovania. [17] Ako už názov napovedá, poskytuje najlepšie hodnotené filmy, ktoré zodpovedajú vkusu používateľa. Je podobný ako PVR s tým rozdielom, že sa pozerá len na hodnotenie titulu. Netflix využíva Tento algoritmus aj na doplnenie riadkov, ako napríklad "Pre teba". [16]

Ďalším algoritmom spoločnosti Netflix je

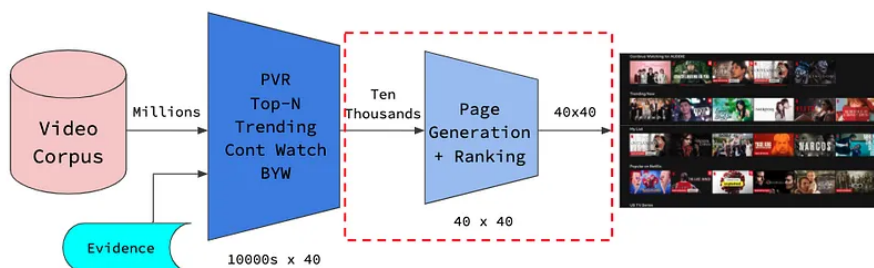
3. **Obľúbene medzi používateľmi**, ktorý predpovedá časové trendy na základe globálnych a regionálnych údajov o filmoch a televíznych seriáloch. Vo všeobecnosti existujú dva typy trendov: sezónne a špeciálne. Na základe názvu by ste mohli predpokladať, že čo je čo, ale poďme sa na ne pozrieť bližšie. [17]

Do sekcie **sezónnych trendov** sa zvyčajne zaraďujú filmy ako Sám doma alebo Polárny expres, pretože získavajú popularitu v období Vianoc, alebo horory počas Halloween. [16]

Sekcia **špeciálne** alebo **krátkodobé** udalosti môže patriť Pandémii 2021, ktorá zaznamenala nárast populárnosti dokumentárnych filmov a filmov o infekciách. [16]

4. **Pozrite sa neskôr**, ide o pomerne zaujímavý algoritmus, ktorý je založený na dvoch kľúčových bodoch. [16] Prvým je, že do tejto časti patria filmy, ktoré ste ešte nedopozerali, zvyčajne televízne seriály s niekoľkými sezónami, ale môžu to byť aj filmy alebo seriály, ktoré ešte neboli dokončené. [18] A druhým bodom je, že na základe regionálnych údajov a pomocou kolaboratívneho filtrovania algoritmus zmieša vaše nepozreté filmy s podobnými filmami, ktoré sa páčia iným používateľom. [17]

Po tom, ako algoritmy analyzujú vás a vaše správanie, začnú vytvárať vaše personalizované odporúčania, krok za krokom prostredníctvom všetkých parametrov, čo je graficky znázornené na obrázku 4. plynulo prechádzajú z backendu na frontend, takýto proces sa nazýva **generovanie stránok**.



Obr. 5: Ranking [16]

6 Typy Údajov v Netflix

A teraz nehovoríme len o tom, aké filmy uprednostňujete, ale o výsledkoch systému, ktorý má viac ako 1 300 skupín odporúčaní skenujúcich vás a vaše zvyky. Netflix vie, že máte radi nielen thrillery; on vie že zaoberáte sa psychologické thrillery so silnými ženskými hrdinkami, ktoré sledujete väčšinou cez víkend večer. Asi vás zaujíma ako. Odpoveď nájdete v tejto časti. [3] [15]

1. **Údaje o účtoch a predplatnom**, na prvý pohľad nie je jasné, prečo Netflix používa tieto údaje, ale je to na lepšiu správu vašich účtov a zaistenie bezpečnosti monitorovaním vašej aktivity na účte. [19]
2. **Zariadenie a dáta**, z ktorého sledujete filmy a v akom čase, aby sú vedeli, čo a kedy vám máme odporučiť. [19] Vezmime si, že v práci počas prestávky dve hodiny pozeráte komediálne seriály a potom skončí a večer si na televízore zapnete Netflix, kde pozeráte horory, zhromažďovaním týchto informácií sa vaše odporúčania začnú meniť podľa dátumu a času.
3. Spoločnosť Netflix zhromažďuje údaje ob **históriu dopytov**, aby mohla sledovať trendy a prispôbovať ponuky obsahu s cieľom zdôrazniť to, čo používatelia hľadajú. [20]

4. Samozrejme, že spracúvajú **hodnotenie obsahu** a **spätnú väzbu** používateľov, aby pochopili kvalitu a hodnotenie obsahu a rozhodli sa, či obsah spracovať alebo odstrániť. [19]

7 Záver

Toto je koniec článku a stojí za to ho zhrnúť. Netflix je rýchlo rastúca platforma, ktorá stojí na vrchole odporúčacieho reťazca vďaka svojim algoritmom a pokročilej umelej inteligencii, ktorá je založená na hlbokom učení. Pomocou rôznych metód filtrovania, umelej inteligencie a systémov hodnotenia, ako sú PVR, Top-N Ranking atď. dokáže vytvoriť pre vás najrelevantnejší a najvhodnejší obsah.

Okrem toho spoločnosť Netflix využíva zhromaždené údaje, ako je zariadenie, čas sledovania a poloha, aby pochopila váš vkus a poskytla vám a ostatným používateľom najvhodnejší obsah. Netflix pracuje aj s rozhraním a poskytuje vám miniatúry, ktoré vytvárajú prvý dojem z filmu alebo seriálu a udržiavajú váš záujem.

Spoločnosť Netflix v súčasnosti skúma problém studeného štartu s cieľom vyriešiť ho. Týmto spôsobom sa spoločnosť Netflix snaží vytvoriť čo najpríjemnejší zážitok pre každého používateľa, aby dosiahla väčší úspech na trhu. Okrem toho pokračuje v štúdiu a zdokonaľovaní umelej inteligencie a techník filtrovania na presnejšiu personalizáciu obsahu.

Literatúra

- [1] Arkadiusz Krysiak. Netflix algorithm: How netflix uses ai to improve personalization. <https://stratoflow.com/how-netflix-recommendation-algorithm-work/>, June 2024.
- [2] Arkadiusz Krysiak. Netflix algorithm: Everything you need to know about the recommendation system of the most popular streaming portal. <https://recostream.com/blog/recommendation-system-netflix>, August 2021.
- [3] James Bennett and Stan Lanning. An introduction to recommender systems at netflix. <https://arxiv.org/abs/1903.04638v7>, 2019.
- [4] Argoid TEam. How netflix uses artificial intelligence. <https://www.argoid.ai/blog/netflix-ai>.
- [5] LITSLINK. Netflix and learn: How netflix uses ai to personalize recommendations. <https://litslink.com/blog/all-about-netflix-artificial-intelligence-the-truth-behind-personalized-content#:~:text=Through%20ML%20algorithms%2C%20Netflix%20continuously,while%20minimizing%20buffering%20and%20interruptions.>, August 2024.
- [6] NETFLIX. Optimizing content quality control at netflix with predictive modeling. <https://netflixtechblog.com/optimizing-content-quality-control-at-netflix-with-predictive-modeling-712281658ab9>, December 2024.

- [7] Puneet Mangla. Netflix movies and series recommendation systems. <https://pyimagesearch.com/2023/07/03/netflix-movies-and-series-recommendation-systems/>, July 2023.
- [8] Yakov Livshits. How netflix’s recommendation algorithm works: A look behind the scenes. <https://habr.com/en/articles/677396/>, July 2022.
- [9] Zhong Hong. Personalized recommendations: How netflix and amazon use deep learning to enhance user experience. <https://medium.com/@zhonghong9998/personalized-recommendations-how-netflix-and-amazon-use-deep-learning-to-enhance-user-experience-e7bd6fcd18ff>, April 2023.
- [10] Prem Kumar. Netflix movie recommendation — using collaborative filtering. <https://towardsdatascience.com/tensorflow-for-recommendation-model-part-1-19f6b6dc207d>.
- [11] Sachin Sharma. The magic of recommendation systems: How netflix knows what you like. <https://www.linkedin.com/pulse/magic-recommendation-systems-how-netflix-knows-what-you-sachin-b9gzc>, February 2024.
- [12] CodeStax.Ai. Creating netflix-worthy recommendations: The collaborative filtering revolution. <https://codestax.medium.com/creating-netflix-worthy-recommendations-the-collaborative-filtering-revolution-a1765c5daec4>, September 2023.
- [13] Shrishail Kandi. Content based movie recommendation system for netflix using natural language processing. <https://shrishailkandi95.medium.com/content-based-movie-recommendation-system-for-netflix-using-natural-language-processing-53fdbcf94417>, June 2021.
- [14] Argoid Team. How does netflix’s recommendation engine work? <https://www.argoid.ai/blog/how-does-netflixs-recommendation-engine-work>., July 2024.
- [15] The Product Folks Team. Unlocking viewer engagement: How netflix uses data to win over 230+ million users worldwide. <https://www.theproductfolks.com/product-management-blog/unlocking-viewer-engagement-how-netflix-uses-data-to-win-over-230-million-users-worldwide>, August 2024.
- [16] Nihal D’Souza. How do netflix recommendations work? <https://nihaldsouza.medium.com/how-do-netflix-recommendations-work-a24385f66d80>, January 2021.
- [17] Arkadiusz Krysik. Inside the netflix algorithm: Ai’s role in personalizing user experience. <https://stratoflow.com/how-netflix-recommendation-syworks/>, August 2024.
- [18] David Chong. Deep dive into netflix’s recommender system. <https://readmedium.com/deep-dive-into-netflixs-recommender-system-341806ae3b48>.

- [19] Lineate. 5 ways netflix uses data to win the world of streaming entertainment. <https://www.lineate.com/blog/5-ways-netflix-uses-data-to-win>, February 2019.
- [20] Kamaljeet Kaur. Behind the screen: How netflix uses data science? <https://www.analyticsvidhya.com/blog/2024/07/netflix-uses-data-science/#:~:text=Netflix%20uses%20its%20data%20to,and%20decide%20on%20content%20investments.>, Aughust 2024.