Sruthi Kapudasi 2024

OPTIMIZING CONTENT RECOMMENDATIONS AT NETFLIX WITH ACTIVE LEARNING

INTRODUCTION:

In today's entertainment landscape, Netflix is successful in being the leading platform for streaming content worldwide. This success is because of its data analytics and recommendation systems tailoring content suggestions to its massive subscriber base of 208 million. It thrives on its data-driven approach to user experience. However, with the ever-growing volume of user data (an astounding 100 billion data points every day, including information on searches, viewing time, content, and abandonment), their data science teams are facing a constant challenge to optimize their recommendation engine to further enhance user satisfaction and retention. Given a team of 1-2, it is challenging to take all the data science concerns into account. This case study helps highlight the different strategies Netflix can employ to navigate these challenges and continue to solidify its position as the go-to destination for entertainment. Even with a growing amount of user data, Netflix's small data science team struggles to keep up with demands like increasing engagement, handling feedback, promoting new content, protecting user privacy, and the filter bubble.

PROPOSED SOLUTION: IMPLEMENTING ACTIVE LEARNING USING THE SCRUM PROJECT MANAGEMENT METHODOLOGY

Active Learning can be used as a key concept for the following issue considering its capability to form clusters to discuss and resolve a course topic. Considering several positive and negative cases for the current issue, active learning explores decision boundaries in an operationally smart approach to gather additional data for training iterations.

- **Solution 1**: Active learning allows the system to pinpoint uncertain recommendations and prioritize user feedback for those specific items. Coming to watch time tracking, to avoid misunderstandings we can track data from the “Continue Watching” area or if it matches previous recommendation lists by considering confounding factors and passive consumption where some users might passively watch without engaging with it in the background. This focused data collection allows the model to learn from user interaction, leading to more accurate and engaging recommendations over time.

- **Solution 2**: Utilize implicit feedback mechanisms to gather data by analyzing watch time patterns (completion rates, drop-off points, pausing, rewinding) to understand user engagement. Also, A/B testing different recommendation layouts or incorporating user reviews alongside suggestions. By using these methods, active learning collects valuable data without burdening users with explicit feedback requests.
- **Solution 3:** Active learning can be implemented by incorporating collaborative filtering techniques that leverage data from similar existing content. We can choose to highlight new content on platforms that match our genre based on past viewing habits and then perform data collection on user preferences and ultimately improve recommendations for both the new content and similar items in the future.
- **Solution 4:** Implement data anonymization techniques to protect user privacy while still allowing the model to learn from user behavior trends. This can happen by prioritizing user feedback on uncertain recommendations which minimizes the overall amount of data collected while maximizing its value for improving model accuracy.
- **Solution 5:** Showcasing trending shows watched by users with similar tastes but slightly different preferences. This follows incorporating exploratory recommendations alongside personalized suggestions where the model can refine its understanding of user preferences and encourage them to discover new and unexpected content instead of being stuck in a filter bubble.

While scrum doesn’t directly address data acquisition, it can be used to manage active learning within the sprint (14 days) as shown below:

S. No	Steps	Time	Reason
1.	Train an initial recommendation model using existing user data	3 days	This model preparation serves as a baseline for identifying areas of uncertainty. This task even aligns with the sprint planning stage in the scrum.
2.	Design mechanisms for users to provide feedback on uncertain recommendations	7 days	This step ensures user feedback is collected efficiently by using Ratings, A/B testing, and implicit feedback through watch time tracking
3.	Model improvement with high uncertainty.	Continuous throughout the sprint	The model retraining process helps analyze and carefully avoid any misinterpretations.
4.	Model Evaluation	3 days	Using a validation set allows us to find areas for further improvement using the newly acquired user feedback data.
5.	Retrospection	1 day	In the end, we discuss team collaboration, successes, challenges, and other areas of improvement.

JUSTIFICATION:

Implementing active learning and user feedback must be an iterative effort and not a one-time effort. As a part of the team, we must make sure to develop a culture of continuous improvement through feedback from the employees who interact with the recommendation system. This looping approach ensures that the learning strategy remains focused on the data quality, which is manageable for a small team, adaptable to user behavior changes, and ultimately helps to improve user engagement by providing the most relevant recommendations.

POSSIBLE CHALLENGES:

The proposed active learning solutions for Netflix's recommendation system offer several advantages, but there are also potential challenges to consider during implementation:

Challenges related to Active Learning –

- Designing Effective Queries - Craft clear and brief questions to get useful feedback for uncertain recommendations. Poorly designed questions lead to user frustration.
- Overfitting and Model Explainability - Active learning can focus too much on specific user data, leading to inaccurate recommendations to wider audiences.

Challenges related to Project Management –

- Limited Resources - A small data team might struggle to manage the process of active learning. Automating tasks and focusing on high-impact areas can help mitigate this.
- Security and Privacy - Transparency with users about data collection practices and offering control over data usage is also important.
- Exploring Alternative Feedback Mechanisms Beyond Explicit Requests - It can be restrictive to only rely on explicit input and lose out on important user insights.

MITIGATION STRATEGIES:

Active learning can fatigue users and overload a small team. Here's how to mitigate that:

- **Streamlined Feedback:** Design user-friendly requests with multiple-choice options or short answers. Use visuals to explain how their input helps with recommendations.
- **Diverse Data Training:** Train the model based on a wider audience instead of just concentrating on uncertainties, this will help in improving the recommendations.
- **Automated Workflows:** Automate repetitive tasks, such as data preparation and collection, model training, and alarms. This allows data scientists to concentrate on main tasks and quickly resolve high-priority problems.

- **Clear Data Practices:** Be transparent about what data is collected for recommendations and provide easy opt-out options for the users. This transparency fosters trust and empowers users.
- **Engaging Review System:** Encourage users to leave reviews periodically, perhaps with gamification elements like points or badges. This fosters a sense of community ownership.

CONCLUSION:

All the challenges and causes identified in the case study are results of making decisions in spate instead of thoroughly understanding and analyzing positive and negative outcomes. Hence, active learning, a reflexive model strategy concept from the textbook "How to Lead in Data Science" written by Jike Chong and Cathy Chang allowed me to understand that it offers a data-driven approach to optimize Netflix's content recommendation system. It equally aligns with project management considerations and goals, allowing a small team to work effectively and address data quality concerns by instantly delivering accurate and relevant recommendations, fostering user engagement and viewership. This approach can solidify Netflix's position as a leader in the streaming industry.

REFERENCES:

1. Concept and solutions for challenges: **How to Lead in Data Science – Book by Jike Chong and Yue Cathy Chang**
2. A general explanation of Active Learning: https://en.wikipedia.org/wiki/Active_learning
3. Active Learning in recommender systems: https://www.researchgate.net/publication/296481444_Active_Learning_in_Recommender_Systems
4. Introduction to Scrum Methodology: **How to Lead in Data Science – Book by Jike Chong and Yue Cathy Chang** and https://resources.scrumalliance.org/Article/scrum-guide-basics