

A case study on Netflix: Data-Driven Decision Making, Assignment-2

To :Dr. Amit Arora

GROUP - 1

Athira Karuvath Raj - 100854112

Dheer Parikh - 100919437

Table of Contents

[INTRODUCTION 2](#_Toc147958141)

[TOOLS AND TECHNOLOGIES USED-HOW THEY ARE USED? 2](#_Toc147958142)

[DATA GENERATION: WHAT IS THE FORMAT DATA IS BEING COLLECTED, STORED OR ANALYSED & DATA STORAGE: WHAT KIND OF STORAGE SYSTEM IS USED? HOW DOES THIS DIFFERENT BETWEEN STRUCTURED AND UNSTRUCTURED DATA? 2](#_Toc147958143)

[DATA COLLECTION & TRANSMISSION: WHAT OVERALL PROCESS WILL YOU USE TO COLLECT THE DATA? WHAT FORMAT WILL THE DATA COME IN? ARE THERE ANY SPECIAL CONSIDERATIONS IN TRANSMITTING THE DATA TO THE STORAGE CENTER? 3](#_Toc147958144)

[DATA ANALYSIS: WHAT KIND OF ANALYSIS DO YOU PLAN TO DO WITH THE DATA? WHAT KIND OF TOOLS MIGHT YOU NEED? WHAT KIND OF VISUALIZATION WILL YOU NEED? 3](#_Toc147958145)

[DECISION MAKING: HOW WILL DECISION MAKERS WANT TO RECEIVE THE DATA?  -DATA GOVERNANCE ISSUES -DATA PRIVACY ISSUES 4](#_Toc147958146)

[REFERENCES: 5](#_Toc147958147)

# INTRODUCTION

We have seen that Netflix's data-driven system, which guides everything from design choices to marketing strategies, is the key to their amazing success. They gambled and invested $100 million in "House of Cards" based solely on analytics that predicted its audience. Since viewers even subscribed just for the show, this risk paid off. Netflix's Original Content division now primarily relies on user data, social media trends, and prediction algorithms to inform its content creation. They are able to make well-informed decisions rapidly because they place a strong emphasis on data processing that is accessible, visually appealing, and efficient. CEO Reed Hastings supports a culture of learning and experimentation and supports large-scale experiments. This data-driven strategy has totally transformed how blockbusters are made and promoted, raising Netflix's value to $140 billion and making Netflix the market leader. In this article we'll explore how Netflix uses a wide range of advanced tools and technologies, such as machine learning techniques, Apache Hadoop, Spark, and visualization tools like Tableau, to implement a comprehensive data-driven approach across all of its operations today.

# TOOLS AND TECHNOLOGIES USED-HOW THEY ARE USED?

Netflix uses a wide range of tools and technologies for data-driven decision-making, including machine learning algorithms like Tree-based Regression and Classification, Support Vector Machine, and many others for analyzing user behavior and preferences. They also use MySQL and NoSQL for data storage, platforms like Apache Hadoop and Spark for large-scale data analysis and processing, and data visualization tools for extracting insights from the data.

# DATA GENERATION: WHAT IS THE FORMAT DATA IS BEING COLLECTED, STORED OR ANALYSED & DATA STORAGE: WHAT KIND OF STORAGE SYSTEM IS USED? HOW DOES THIS DIFFERENT BETWEEN STRUCTURED AND UNSTRUCTURED DATA?

When a user interacts with the platform, the data is collected in real-time and securely transmitted. Mainly the data collected from the user is through.

1) Responsiveness to shows/movies

2) Date, time, location & the device being used to watch

3) Pausing and resuming time.

4) Shows completed / leave midway

5) Time in minutes / hours / days / weeks took to complete a series / movie

6) Number of times a user searching before choosing the show / movie

7) Queries user use to search their shows / movies

8) Shows preferred by different categories such as men / women / children / teenagers

9) Feedback & ratings of subscribers

10) Scrolling behaviour of user

https://www.engati.com/blog/predictive-analytics

The data is formatted in both structured (e.g., JSON, XML) in MySQL and unstructured formats (e.g., text feedback, comments) in NoSQL such as Cassandra, Hive, IBM Cloudant, MongoDB, etc. and in terms of security, Industry-standard security protocols like TLS/SSL are employed to safeguard data against unauthorized access.

# DATA COLLECTION & TRANSMISSION: WHAT OVERALL PROCESS WILL YOU USE TO COLLECT THE DATA? WHAT FORMAT WILL THE DATA COME IN? ARE THERE ANY SPECIAL CONSIDERATIONS IN TRANSMITTING THE DATA TO THE STORAGE CENTER?

Netflix fetch the information from users through various activities on the platform such as viewing history, ratings and reviews, search queries, and user feedback. In simple terms as users interact with the platform, data is captured in real time. The user’s data is then securely transported to their data centers for storage and analysis1 and in order to safeguard user privacy, the data will most likely be transmitted using strong security measures as mentioned above.

For transporting data to storage, Netflix relies on the Cloud for all of its scalability, compute, and storage requirements. Through fibers, data can practically travel at the speed of light so in order to assure zero lag and rapid speeds, hyperscale data centers that hold massive amounts of data are coupled with ultra-high fiber count cable, with up to 10K fibers traveling through ducts.

# DATA ANALYSIS: WHAT KIND OF ANALYSIS DO YOU PLAN TO DO WITH THE DATA? WHAT KIND OF TOOLS MIGHT YOU NEED? WHAT KIND OF VISUALIZATION WILL YOU NEED?

Netflix has emerged as a pioneer in the dynamic environment of digital entertainment, but it is not the only way of its massive content creation but also for its unique approach to recognizing and responding to customer tastes. The tools used for the data analysis are Advanced Analytics where the Netflix employs complex analytics techniques to uncover trends, correlations, and patterns in user behavior which includes analyzing viewing history, ratings, search queries, and user feedback. For prediction and forecasting, Netflix employs various machine learning algorithms such as Tree Based Regression, Support Vector machines and many more such algorithms by selected specific hyperparameters to analyze user behavior and preferences. In order to roll an update Netflix do Experimentation and A/B testing methodologies, which are used to validate hypotheses and optimize the customer experience. This includes testing several versions of a feature with different user segments and comparing the results.

DATA PRE-PROCESSING: WHAT PRE-PROCESSING WILL YOU NEED TO DO ON THE DATA? INTEGRATION? CLEANING? ELIMINATION OF REDUNDANT DATA? ERROR CORRECTION? TRY TO BE AS SPECIFIC AS POSSIBLE. WHAT TYPE OF SECURITY AND PRIVACY CONCERNS ARISE WITH DATA TRANSMISSION?

As we have seen in previous question how Netflix integrate data collected from user and transmit it using strong security measures to their cloud storage, lets understand about why Netflix needs data cleaning. They clean the data to remove any inconsistencies and errors which ensures that data used is accurate and reliable. In addition to that they also remove duplicate data records which helps in reducing the processing time and improve efficiency while doing analysis, this process is known as Elimination of Redundant Data. In addition to this Netflix also performs Error Correction using various algorithms to ensure data quality and integrity. This could involve techniques like outlier detection and removal, filling in missing values, and correcting erroneous values.

# DECISION MAKING: HOW WILL DECISION MAKERS WANT TO RECEIVE THE DATA?  -DATA GOVERNANCE ISSUES -DATA PRIVACY ISSUES

When comes to decision making, the decision maker in Netflix receive data in an actionable and understandable format. For instance, User interaction, content performance, and subscriber growth are all covered in reports with the help of tools such as tableau and power BI.

In terms of Data Governance, Netflix is likely to have a thorough data governance system in place to assure the quality, consistency, and security of its data. They set a limit to who has access to sensitive information, authentication mechanisms to verify user identities, audit logs to track data access and modifications, transparency and accountability, and regular assessments to identify and mitigate potential risks.

In terms of Protecting user privacy, it is a top priority for Netflix. They are most likely using strong security measures to protect user data during collection, transfer, storage, and analysis. TLS encryption of customer information, search queries, and other confidential data is one of these measures. They use pre-encoded Digital Rights Management (DRM) to secure their video broadcasts.

# REFERENCES:

1 <https://research.netflix.com/research-area/machine-learning>

2 <http://netflixtechblog.com/>

3 <https://www.mage.ai/blog/netflix-use-ml-to-become-worlds-streaming-leader>

4 <https://shirshadatta2000.medium.com/>

5 <https://outsideinsight.com/>

6 <https://corpgov.law.harvard.edu/>

7 <https://www.techtarget.com/searchsecurity/feature/10-types-of-security-incidents-and-how-to-handle-them>