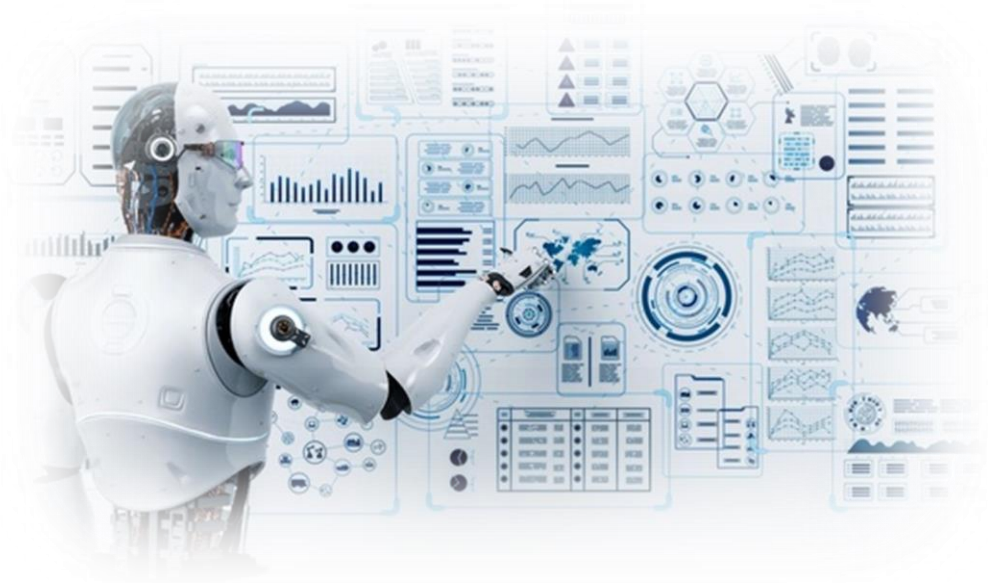# CONTENT RECOMMENDATION SYSTEM

Project Report

Submitted On :14-01-2024

Submitted By:

Tushar Sehdev
Museib Patel
Mohd Abdul Ghani
Dhatchana Moorthy Mathan

Submitted To:
F13 Technologies

Contents

Introduction

Content recommendation system is a useful tool that can improve user experience, engagement, and conversion by offering personalized and relevant suggestions based on user preferences and behaviour. Content recommendation system can be applied to various types of content, such as products, articles, videos, and more, across different industries, such as e-commerce, media, and entertainment. Content recommendation system can help businesses differentiate themselves from competitors, increase customer loyalty, and maximize revenue.

AWS Personalize is a fully managed machine learning service that allows developers to easily build and deploy custom content recommendation system without any prior machine learning knowledge. AWS Personalize uses advanced algorithms, such as collaborative filtering and personalized ranking, to analyse user data and generate recommendations for different use cases, such as similar items, user segmentation, and unstructured text. AWS Personalize also supports real-time updates and A/B testing to improve the recommendation quality and user satisfaction.

In an era where data is abundant and user expectations are ever-evolving, the CRS is poised to be a game-changer, ushering in a new era of tailored, user-centric service recommendations. This report serves as a roadmap to navigate the depths of this innovative recommendation system, ultimately unlocking its potential to transform the way we interact with digital services.

CONTENT RECOMMENDATION

Data Processing:

- During the data processing phase, information from the Movie Lens dataset is extracted and structured into three separate files: links, movies, and ratings.

- The links file typically serves as a repository for linking movie IDs to external databases or sources, aiding in cross-referencing and obtaining additional contextual information.

- The movies file contains metadata for each movie, encompassing details like title, genre, and release year, essential for content identification and classification.

- The ratings file contains user-generated ratings for specific movies, reflecting user preferences and engagements with the content. This organized data framework facilitates streamlined analysis and modeling, supporting tasks such as developing recommendation systems and comprehending user behavior effectively.

	A	B	C
1	MOVIE_ID	IMDB_ID	TMDB_ID
2	1	114709	862
3	2	113497	8844
4	3	113228	15602
5	4	114885	31357
6	5	113041	11862
7	6	113277	949
8	7	114319	11860
9	8	112302	45325
10	9	114576	9091

	A	B	C	D	E	F
1	ITEM_ID	TITLE	GENRES			
2	1	Toy Story	Adventure|Animation|Children|Come
3	2	Jumanji (1	Adventure|Children|Fantasy
4	3	Grumpier	Comedy|Romance
5	4	Waiting to	Comedy|Drama|Romance
6	5	Father of	Comedy
7	6	Heat (199!	Action|Crime|Thriller
8	7	Sabrina (1	Comedy|Romance
9	8	Tom and F	Adventure|Children
10	9	Sudden D	Action
11	10	GoldenEy	Action|Adventure|Thriller

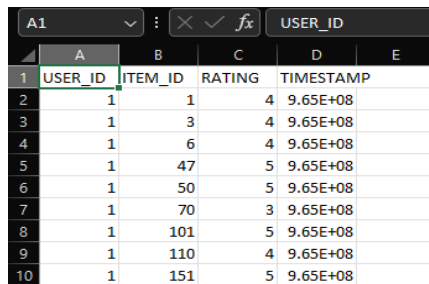|	A	B	C	D
1	USER_ID	ITEM_ID	RATING	TIMESTAMP
2	1	1	4	9.65E+08
3	1	3	4	9.65E+08
4	1	6	4	9.65E+08
5	1	47	5	9.65E+08
6	1	50	5	9.65E+08
7	1	70	3	9.65E+08
8	1	101	5	9.65E+08
9	1	110	4	9.65E+08
10	1	151	5	9.65E+08
11	1	157	5	9.65E+08

Image i: Link file
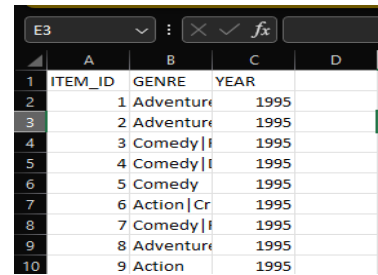*Image ii: movies file*
Image iii: ratings files

Data Pre-Processing:

The first step in building the CRS was to collect and prepare the data. A sample dataset was created, containing files item-meta data & interactions where data of USER_ID, ITEM_ID, TIMESTAMP, EVENT_TYPE are filled. This data is pre processed from the movies lens data files using python. So it can be suitable for deployment and import files.

Image 1 Loading dataset in excel interactions files
Image 2 Loading dataset in excel item-meta data files

Data Processing: This dataset was then uploaded to AWS S3 for storage. And then connect that S3 bucket movie-rs to AWS Personalize for CSR.
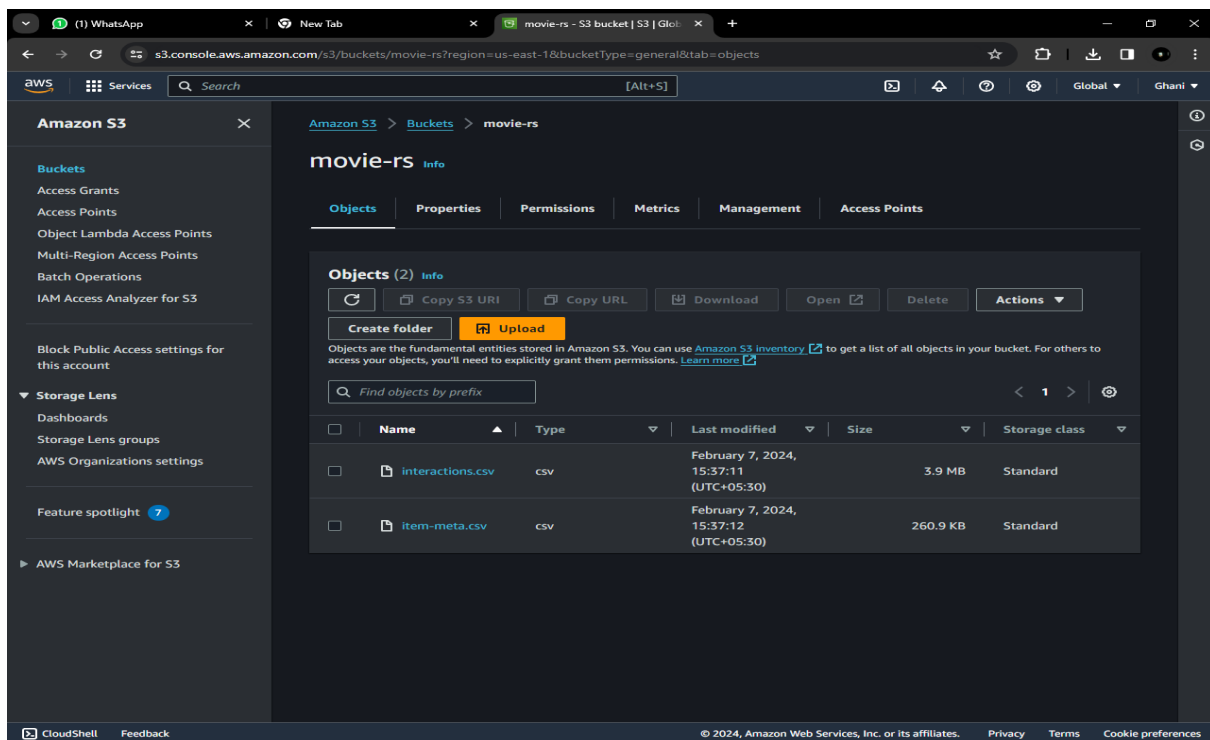

Image 3 Importing all the required libraries in movies-rs.

Aws Personalize:

Now we create a Data set in AWS Personalize, here first step we create a Dataset group with name "movies_rs" in which we import the dataset in .csv format from s3 bucket of movies-rs.
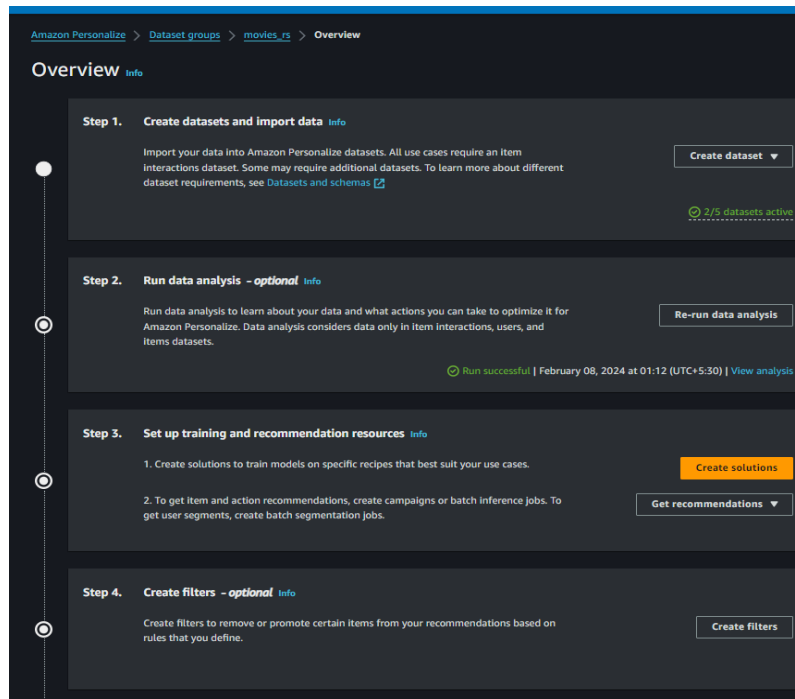

Image 4 Dataset is created and imported from s3 bucket

Solution & Recipe:

In this phase, we generate solutions by selecting from a set of predefined recipes available in AWS. We choose the appropriate recipe for each of the three solutions we have created.
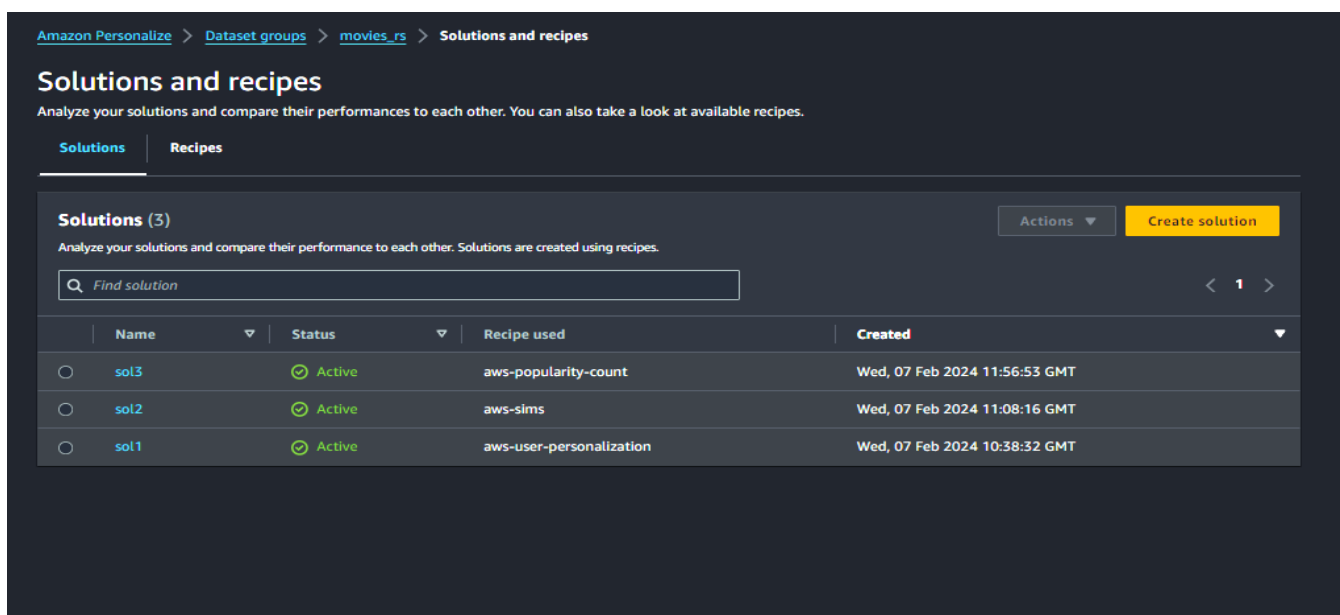

Image 5 Solutions and recipes

Runing Campaign:

In this phase we create a campaign to the solution, configure settings such as campaign budget, duration, target audience, geographic targeting, etc., depending on your campaign requirements. Start the campaign by deploying your resources and activating your advertising or promotional content. Make necessary adjustments to your campaign settings or resources to optimize performance based on real-time data analysis.

Image6:Campaign

Testing:

Now set the target audience parameters. Then, deploy resources, activate promotional content, and monitor performance metrics to optimize results effectively.

Item ID	Score	GENRE	YEAR
4246	0.2398668	Comedy|Drama|Romance	2001
33679	0.0494432	Action|Adventure|Comedy|Romance	2005
30812	0.0404081	Drama	2004
2671	0.0341613	Comedy|Romance	1999
5299	0.0259780	Comedy|Romance	2002
5991	0.0238597	Comedy|Crime|Drama|Musical	2002
5902	0.0160457	Comedy|Drama|Romance	2002
5377	0.0135400	Comedy|Drama|Romance	2002
8972	0.0111186	Action|Adventure|Drama|Mystery|Thriller	2004
40629	0.0075100	Drama|Romance	2005
5945	0.0067030	Comedy|Drama	2002
7445	0.0064314	Action|Crime|Drama|Mystery|Thriller	2004
8798	0.0062158	Action|Crime|Drama|Thriller	2004
8638	0.0061863	Drama|Romance	2004
8970	0.0061678	Drama	2004
45517	0.0059847	Animation|Children|Comedy	2006
5459	0.0058783	Action|Comedy|Sci-Fi	2002
91500	0.0058630	Action|Adventure|Drama|Sci-Fi|Thriller	2012
4776	0.0058177	Crime|Drama|Thriller	2001
5013	0.0055021	Comedy|Drama|Mystery	2001
6947	0.0054700	Adventure|Drama|War	2003
34162	0.0051731	Comedy|Romance	2005
6873	0.0051277	Comedy|Romance	2003
5954	0.0049801	Crime|Drama	2002
4447	0.0049439	Comedy|Romance	2001

Image7: Campaign running successfully

Website Development:

We have created website development process, emphasizing the use of Git for version control and incorporating Node.js modules and CSS for website styling.
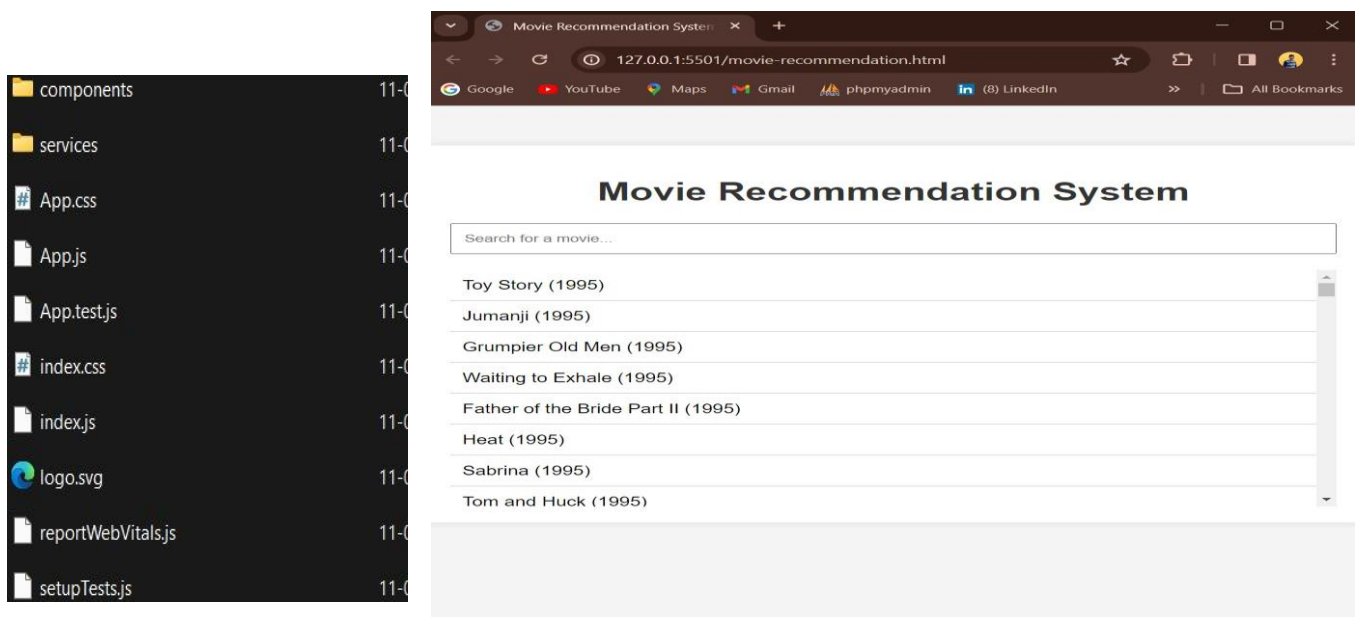
Image8: Website and tools used

7

Preference and Behavior:

User preferences and behavior are captured implicitly through the system's interactions. When a user selects or engages with a service, the CRS can record this behavior and use it to refine recommendations. However, this report does not cover the explicit mechanisms for collecting user data.

The camp1: Precision At 5 0.0679 At 10 0.0482 At 25 0.0314
The camp2: Precision At 5 0.0676 At 10 0.0426 At 25 0.0259
The camp3: Precision At 5 0.0116 At 10 0.0116 At 25 0.0099
In this we see precision speed of camp1 is high so we are using solution1 for our website.

Architecture Diagram

```
+------------------+
||                ||
||   User Data    ||
||  (Interactions)||
||   Amazon S3    ||
+--------+---------+
         |
         v
+------------------+
||                ||
||   AWS IAM      ||
|| (Identity & Access ||
||   Management)  ||
+--------+---------+
         |
         v
+------------------+
||                ||
||  AWS Personalize ||
||                ||
+--------+---------+
        ---|
           v
+------------------+
||                ||
||  Personalized  ||
||  Recommendations ||
||                ||
+--------+---------+
         |
         v
+------------------+
||                ||
||  User Interface ||
||   Application  ||
||                ||
+--------+---------+
```

Cost Analysis

Estimate URL: **https://calculator.aws/#/estimate?
nc2=h_ql_pr_calc&id=275d8bfdc49dca4e44a91d2f88e4d1b394935f90**

Estimate summary

Upfront cost	Monthly cost	Total 12 months cost
0.00 USD	128.12 USD	1,537.44 USD
		Includes upfront cost

Detailed Estimate

Name	Group	Region	Upfront cost	Monthly cost
Amazon Personalize	No group applied	US East (N. Virginia)	0.00 USD	125.73 USD

Status: -
Description: Content Recommendation System
Config summary: Average amount of data ingested per month (10 GB per month), Number of Users in dataset (500), Number of hours recommender is active per month (200), Number of additional recommendations per hour (25), Number of hours with additional recommendations per month (50), Average amount of data ingested per month (8 GB per month), Average training hours per month (30), Number of Users in dataset (500), Average amount of data ingested per month (10 GB per month), Average training hours per month (50), TPS-hours per month for real time inference (150), Number of batch recommendations per month (80)

Name	Group	Region	Upfront cost	Monthly cost
Amazon Simple Storage Service (S3)	No group applied	US East (N. Virginia)	0.00 USD	2.39 USD

Status: -
Description: Cost of uploading data in s3
Config summary: S3 Standard storage (100 GB per month), PUT, COPY, POST, LIST requests to S3 Standard (500), GET, SELECT, and all other requests from S3 Standard (1000), Data returned by S3 Select (50 GB per month), Data scanned by S3 Select (25 GB per month) DT Inbound: All other regions (5 TB per month), DT Outbound: Amazon CloudFront (10 TB per month)

Acknowledgement

Performance Analysis

Cloud Watch Logs Insights
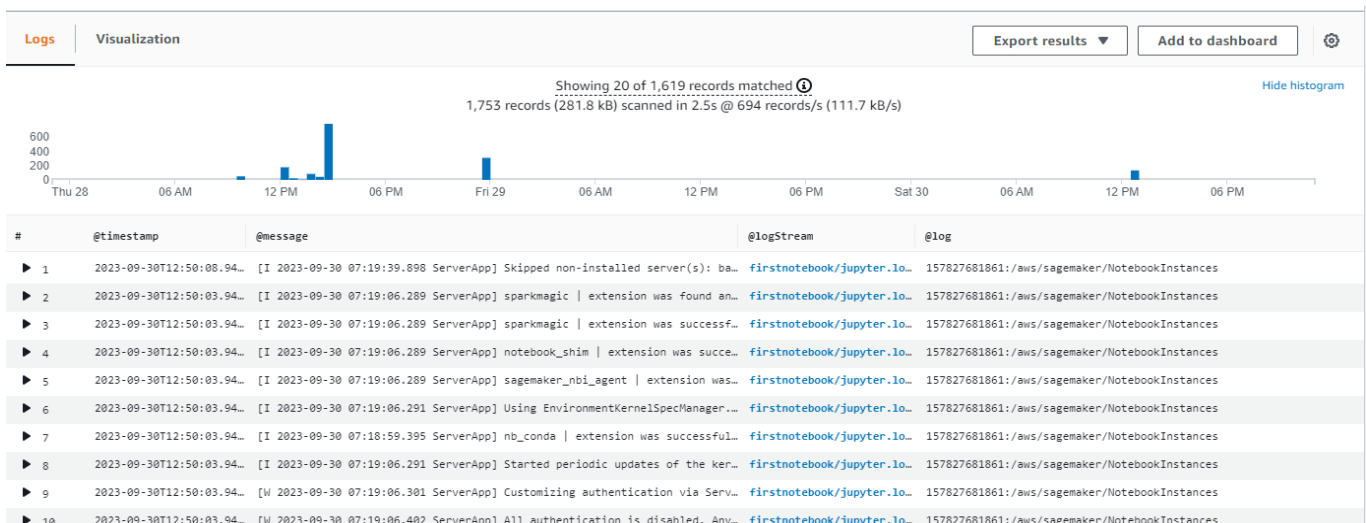
Aws services Used

Amazon S3 (Simple Storage Service):

Amazon S3 is an object storage service provided by Amazon Web Services (AWS). It is designed to store and retrieve vast amounts of data securely and cost-effectively. S3 offers scalable, durable, and highly available storage, making it a popular choice for storing various types of data, including files, images, videos, and datasets.

Key Features of Amazon S3:

Object-Based Storage: S3 stores data as objects, each consisting of data, a unique key, and metadata.

Scalability: It scales seamlessly to accommodate any volume of data, from gigabytes to petabytes.

Durability: S3 redundantly stores data across multiple data centers, ensuring 99.999999999% durability.

Security: S3 provides access control mechanisms and supports encryption to protect data.

Data Lifecycle Management: You can automate data retention and deletion policies.

Data Versioning: S3 allows versioning of objects, enabling data recovery in case of accidental deletions or overwrites.

Amazon Route 53:

Amazon Route 53 is an AWS-provided scalable Domain Name System (DNS) service, translating domain names into IP addresses for efficient internet connectivity.

Key Features:

1. DNS Translation: Translates domain names into IP addresses for seamless connection to internet resources.

2. Scalability: Handles high volumes of DNS queries and traffic, ensuring reliable performance.

3. High Availability: Operates across multiple geographically dispersed data centers for robust and highly available DNS resolution.

4. Health Checking and Failover: Monitors application endpoints' health and automatically routes traffic to healthy resources.

5. Traffic Flow Management: Allows control over traffic routing based on parameters like geographic location, latency, and endpoint health.

6. Security: CloudFront provides security features such as HTTPS support, AWS Web Application Firewall (WAF) integration, and DDoS protection.

7.Cost Efficiency: CloudFront helps reduce data transfer costs by caching content and reducing the load on origin servers.

8.Origin Fetching: It can fetch content from various origins, including Amazon S3, EC2 instances, and on-premises servers.

Other AWS Services (Optional)

Amazon EC2 (Elastic Compute Cloud):

Amazon EC2 is a core AWS service that provides resizable compute capacity in the cloud. It allows you to create and manage virtual servers known as EC2 instances. Here are key points about EC2:

a) **Elasticity:** EC2 instances can be easily scaled up or down to accommodate changing workloads, making them suitable for hosting applications of all sizes.

b) **Customization:** Users have full control over the operating system, software, and configurations of EC2 instances, allowing for highly customized hosting environments.

c) **Security**: Security features like security groups, network ACLs, and IAM (Identity and Access Management) allow users to secure their instances and applications.

IAM User (Identity and Access Management User):

An IAM User is a persona within AWS Identity and Access Management (IAM) that represents an individual entity or application requiring access to AWS resources.

a) Access Control: IAM Users are assigned permissions to access specific AWS resources, allowing fine-grained control over actions they can perform.

b) Multi-Factor Authentication (MFA): IAM supports MFA, adding an extra layer of security by requiring additional verification beyond just a username and password.

c)Access Keys: IAM Users can generate access keys for programmatic access to AWS services via APIs, enhancing automation capabilities.

Amazon CloudWatch:

Amazon CloudWatch in a personalized recommendation system implemented using AWS Personalize, operators can effectively monitor, troubleshoot, and optimize the system's performance and reliability, ensuring a seamless and personalized user experience.

In-Depth Analysis of SKY Content Recommendation System

Introduction:

Content Recommendation Systems (CRS) have become an integral part of many online platforms, aiding users in discovering relevant content in the vast sea of options. In this report, we will delve into the workings of an existing CRS by examining Sky's recommendation system. We will analyze its features, architecture, algorithms employed, and overall effectiveness.

Sky's recommendation system overview:

Sky, a European media and telecommunications company, employs a sophisticated recommendation system powered by AWS Personalize to enhance user engagement and satisfaction across its platforms.

a. Personalization: Sky's content recommendation system tailors' suggestions for individual users by analyzing their viewing history, preferences, and interactions across the platform.

b. Diverse Recommendation Types: Sky provides recommendations spanning various content categories such as TV shows, movies, documentaries, and sports events, catering to a broad audience.

c. Discovery and Surprise: The system balances familiar suggestions with unexpected content choices, encouraging users to explore new shows and genres.

d. Individual Profiles: Sky allows multiple user profiles within a single account, with each profile receiving personalized recommendations based on its unique viewing patterns.

Architecture:

a) Sky's content recommendation system architecture likely involves multiple components:

b) Data Ingestion: Integration with various data sources such as user interaction data, content metadata, and contextual information.

c)
Data Processing: Pre-processing of raw data to extract features, transform data into suitable formats, and perform data cleaning tasks.

d)
Model Training: Utilization of AWS Personalizes machine learning algorithms to train recommendation models based on historical user behavior and content attributes.

e)
Model Deployment: Deployment of trained models to generate real-time

recommendations for users across Sky's platforms.

f)

Feedback Loop: Integration of user feedback mechanisms to continuously update and refine recommendation models.

The architecture may also include additional components for monitoring, logging, and scalability.
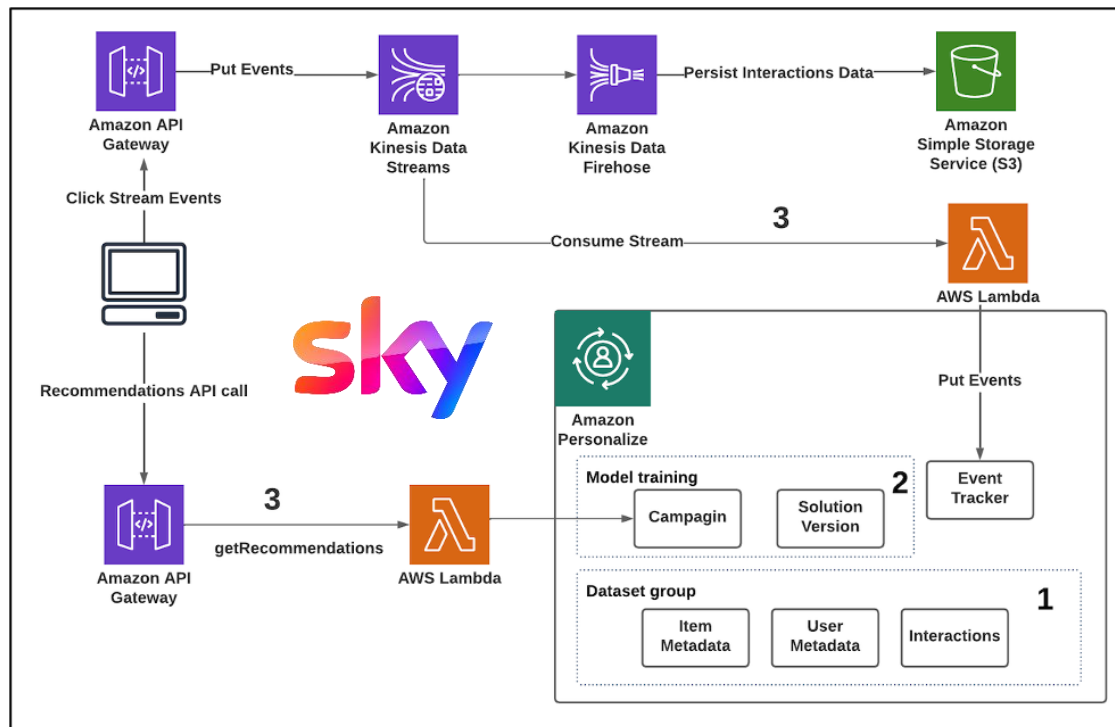

Image 6 Sky Recommendation system Architecture

Algorithms Employed:

Sky's CRS utilizes several recommendation algorithms, including:

a. Sky likely employs a variety of machine learning algorithms provided by AWS Personalize, such as:

b. Collaborative Filtering: Analyzing user interactions and similarities to recommend content based on similar user preferences.

c. User Personalization: Building personalized recommendation models tailored to individual user preferences and behavior.

d. Item-to-Item Recommendations: Recommending similar items based on content attributes and user preferences.

e. These algorithms are trained using historical data and optimized to provide accurate and relevant recommendations to Sky's users.

16

f. Contextual Bandits: To consider temporal and situational context in recommendations.

Overall Effectiveness:

Sky's recommendation system is highly effective and plays a pivotal role in retaining subscribers. It is estimated that over 80% of the content watched on Sky's is driven by recommendations. The system's success can be attributed to its ability to:

a. Understand User Behavior: Sky comprehensively analyzes user behavior, including viewing history, time of day, and device used.

b. Adapt in Real-Time: Recommendations are updated frequently to reflect user interactions, leading to continuous improvement.

c. Balance Exploration and Exploitation: The system encourages users to explore new content while also catering to their known preferences.

d. Handle Cold Start Problem: Sky effectively addresses the "cold start" problem for new users by leveraging content-based recommendations initially.

Conclusion:

Sky's content recommendation system serves as an exemplary case study of a hybrid recommendation system's capabilities. Its personalized, real-time, and adaptable nature makes it a powerful tool for content discovery, contributing significantly to user satisfaction and platform success. The combination of collaborative filtering, content-based filtering, and machine learning algorithms forms the backbone of this highly effective system, providing valuable insights for the field of recommendation systems.

Conclusion

The Content-Based Recommendation System, utilizes natural language processing and cosine similarity to provide service recommendations to users. By processing textual descriptions of services and calculating their similarity scores, the system can generate relevant recommendations. User preferences and behavior can be leveraged to further enhance the recommendation quality and personalization.

This report has provided an overview of the CRS's working, from data preparation to recommendation generation. It serves as a foundation for understanding the system's functionality and potential for future improvements.

References

- Amazon.com
- YouTube
- LinkedIn

Tools Used

- ChatGPT
- MS Office
- VS. Code
- Adobe Acrobat
- MS Excel