Navigating the Anime Universe

The Creation and Implementation of Otaku Castle's Recommender System

Task 2

Joshua C. Rogers

Western Governors University

Table of Contents

Introduction

In the rapidly expanding world of anime, enthusiasts are often inundated with a plethora of choices. From age-old classics to contemporary masterpieces, the anime universe offers a rich tapestry of stories, genres, and art forms. While this vastness presents opportunities for exploration, it simultaneously poses challenges, especially for newcomers seeking their next watch or seasoned viewers looking to discover hidden gems. Otaku Castle, recognizing this challenge, identified an opportunity to enhance its platform by offering tailored anime recommendations to its users.

Modern technology, fueled by advancements in machine learning, has enabled platforms to provide personalized experiences to their users. By understanding user preferences and behaviors, these systems can suggest content that aligns closely with individual tastes. In the context of anime, where the sheer volume of available content can be overwhelming, a recommendation system can streamline choices and create a more engaging user experience.

This paper delves into the intricacies of designing and implementing such a recommendation system for Otaku Castle. By harnessing the power of supervised machine learning and utilizing the Neural Collaborative Filtering model, the goal is to transform the user experience, positioning Otaku Castle as a trusted guide in the vast anime universe.

A. Proposal Overview

A.1 Organizational Need

The company Otaku Castle would like to increase interaction with visitors to its website. One way to do that is to create an anime recommender system. This system would use supervised machine learning to let visitors/customers select a few anime they have watched and give those a rating. The machine learning algorithm would then select and present several other anime the visitor might be interested in watching based on their initial selections and ratings.

A.2 Context and Background

Otakucastle.com is an anime blog/store. Both the site and its visitors would benefit from having a page on the site dedicated to an anime recommendation system. By providing its customers recommendations on new anime to watch based on what they have previously watched compared with ratings from other users, the site would gain more traffic. This in turn would lead to higher revenue for the site.

A.3 and A3A Summary of Published Works and Their Relation to the Project

Review of Work 1

The Nvidia Recommender System web page is a wealth of information. It describes the various methodologies and techniques used in building a recommender system, from matrix factorization to deep neural network models. The page gives an overviw of several available methods, broken down by the 2 main types of systems: Collaborative Filtering and Content and Context Filtering. The page goes on to cover current DL-based models including DLRM, Wide

and Deep (W&D), Neural Collaborative Filtering (NCF), Variational AutoEncoder (VAE), and BERT, among many other methods.

The section on Collaborative Filtering, which is the type of system I chose, states that these systems ”recommend items (this is the filtering part) based on preference information from many users (this is the collaborative part). This approach uses similarity of user preference behavior, given previous interactions between users and items, recommender algorithms learn to predict future interaction. […] For example, if a collaborative filtering recommender knows you and another user share similar tastes in movies, it might recommend a movie to you that it knows this other user already likes.” (Nvidia, 2023)

Further within the web page is a section called “Why Recommendation Systems Run Better with GPUs” discusses briefly that most recommenders, as a form of matrix mathematical learning systems, perform significantly better on one or more GPUs. “The mathematical operations underlying many machine learning algorithms are often matrix multiplications. These types of operations are highly parallelizable and can be greatly accelerated using a GPU. A GPU is composed of hundreds of cores that can handle thousands of threads in parallel. Because neural nets are created from large numbers of identical neurons they are highly parallel by nature. This parallelism maps naturally to GPUs, which can deliver a 10X higher performance than CPU-only platforms.” (Nvidia, 2023) I found this to be true with my own system while creating the model on the CPU versus GPU. Training time was overwhelmingly better when run on the GPU. There is a more in-depth discussion of this in section C3.

Neural Collaborative Filtering fulfils the organizational need by allowing a new user to the site to receive recommendations on what anime to watch next. The results provided to the

new user are based on ratings the user gives to a selection of anime and takes those scores into account when the model runs, looking for similar users ratings.

Review of Work 2

The second source I used comes from FreeCodeCamp.org, a fantastic resource for anyone learning just about any aspect of programming. The article itself, titled "How companies use collaborative filtering to learn exactly what you want" provides a good overview on what collaborative filtering is and how it works. It explains that companies like Amazon and Netflix use collaborative filtering to make product and content recommendations.

This type of system predicts a user's interests by collecting preferences from many users. "Collaborative filtering works on a fundamental principle: you are likely to like what someone similar to you likes. The algorithm's job is to find someone who has buying or watching habits similar to yours, and suggest to you what he/she gave a high rating to." (Arasanipalai, 2019)

The main idea is that if two users agree on one issue, they will likely agree on others as well. However, there are challenges, such as the "cold start problem" when there's little data on new users or products. "Collaborative filtering works well when you have two things:

- a lot of data on what each customer likes (based on what they previously rated high)
- a lot of data on what audience each movie or product might cater to (based on the type of people who rated it high).

But how about new users and new products, for which you don't have much information? Collaborative filtering doesn't work well in these scenarios, so you might have to try something else. Some common solutions involve analyzing metadata or making new users go through a few questions to learn their initial preferences." (Arasanipalai, 2019)

The Nvidia article detailed the problem solved by the design of this recommender system. Otaku Castle is preparing the recommendation system for new users. There is no data on what anime the new users have seen or liked. The first page a user sees when asking for recommendations about what to watch next provides the user with 10 choices from the top 30 rated anime. Once they have selected a rating for each, on a scale from 1 - 10 plus "Haven't seen it", the algorithm is able to make good recommendations for the new user.

Review of Work 3

The third article, titled "Neural Collaborative Filtering: Supercharging Collaborative Filtering with Neural Networks" comes from Towards Data Science, an excellent resource for new and experienced data scientists. I have followed their work for some time, and learned through articles and classes. In this article, they explain that with the glut of information, recommender systems are crucial for curbing information overload.

These systems predominantly use Collaborative Filtering (CF) with implicit feedback. Traditional CF employs matrix factorization to learn user-item interactions. However, Neural Collaborative Filtering (NCF) enhances this by replacing the user-item inner product with a neural architecture. As stated in the article, "Despite the effectiveness of matrix factorization for collaborative filtering, it's performance is hindered by the simple choice of interaction function - inner product. Its performance can be improved by incorporating user-item bias terms into the interactiion function. This proves that the simple multiplication of latent features (inner product), may not be sufficient to capture the complex structure of user interaction data.

This calls for designing a better, dedicated interaction function for modeling the latent feature interaction between users and items. Neural Collaborative Filtering (NCF) aims to solve this by:

1) Modeling user-item feature interaction through neural network architecture. It utilizes a Multi-Layer Perceptron(MLP) to learn user-item interactions. This is an upgrade over MF as MLP can (theoretically) learn any continuous function and has high level of nonlinearities(due to multiple layers) making it well endowed to learn user-item interaction function." (Sharma, 2019)

By using NCF, Otaku Castle will be able to make better recommendataions. Simpler solutions for supervised machine learning, such as those found within Scikit-Learn, do not provide the desired results for their new users. The type and scope of anime and user data available from Kaggle is better suited to an NCF solution for a recommender system.

A.4 Summary of Data Analytics Solution

Otaku Castle is a growing spot for anime fans and wants to make its website even better for its visitors. Knowing that it can be hard for people to pick from so many anime options, they had an idea: why not create a system that suggests anime? By using a neural network style of machine learning, this system lets visitors rate a few anime they've seen. From there, the model processes their ratings then gives them a list of five other anime they might like based on their ratings.

Analytic Method and Implementation (from B3 and C4):

- Analytic Method: Leveraging the power of Neural Collaborative Filtering (NCF), a cutting-edge approach tailored for recommendation systems. NCF combines matrix factorization and deep learning, providing a comprehensive understanding of user preferences.
- Implementation Process: Planning: Prioritize the establishment of a foundational infrastructure for the NCF model and create an interactive web page for user input.
- Sprint Execution: Embark on tasks such as data preprocessing, model training, and user interface design for the web page.
- Testing and Feedback: Post each sprint, the developed functionalities undergo rigorous testing, ensuring seamless user experience and accurate recommendations.
- Release and Maintenance: After achieving a satisfactory set of features, the recommendation system is launched, allowing users to immerse themselves in a personalized anime journey. Continuous monitoring ensures sustained performance and integration of user feedback for enhancements.

Metrics and Benchmark (from C4):

- Metric: Loss, a tangible measure of the gap between the model's predictions and actual user ratings.
- Benchmark: The goal is to achieve a training loss nearing 1.6, ensuring that the validation loss remains in close tandem. This ensures precise predictions and the model's adeptness in catering to new, unseen data.

The solution designed fits well with what Otaku Castle wants. Using the Neural Collaborative Filtering (NCF) model helps give users anime suggestions that they'll likely enjoy. Our step-by-step approach to building and testing the system makes it easy to put into action and adjust as needed. The tools and methods we've chosen are practical, and they make sure the system works well. With this system in place, Otaku Castle can become a favorite spot for anime fans to get personalized recommendations. In short, our solution is both doable and matches the goals of Otaku Castle, aiming to boost website visits and keep users coming back.

A.5 Benefits and Support of Decision-Making Process

Benefits:

1. New Content Discovery
2. Personalized Recommendations
3. Increased User Engagement
4. Data-Driven Insights

With the large amount of anime available for viewing, visitors to Otaku Castle might be overwhelmed when trying to find a “new to them” series or movie to watch. By providing a way for new users to rate anime they have seen the NCF model is able to provide personalized recommendations. These will get better over time, as the user rates additional anime. Further, the integration of a recommender system into the website will provide increased user interaction, a key element in user retention and brand loyalty. With users spending more time on the site, there is also an increased likelihood of making purchases related to a recommendation made by the model. Finally, the site will be given data-driven insights into user behavior, preferences, and trends, which can be leveraged for marketing and other strategic decisions.

Decision Making Process:

1. Strategic Partnerships
2. Marketing and Promotions
3. User Experience Enhancements
4. Monetization Opportunities

Otaku Castle has no intention of becoming a steaming service for anime, however by directing users to external streaming platforms, it positions itself as a trusted recommendation site, providing value to both users and potential strategic partnerships. The data that can be collected with new users recommendations and tracking which anime they watch (at a streaming partner) would provide opportunities to market products to the users directly tied to anime they rate the highest. This data also would apply to promotional email campaigns and targeted advertising, leveraging the users ratings to attract additional new users to the platform. This same data would allow for future site modifications that enhance a users experience, once again driving new and repeat users to the platform. Finally, the data this model provides can lead to new monetization opportunities that have not even been thought of at this juncture.

B. Data Analytics Project Plan

B.1 Goals, Objectives, and Deliverables

- The goal of this project is to provide users of the website www.otakucastle.com a page where they can select anime they have seen, give those anime a rating of their own and they will be given a list of five anime they may be interested in watching next.
- The objective is to create a machine learning system that takes a dataframe of anime and the user rated scores and then is able to output several anime a user may be interested in watching next based on scores they give to a selection of anime presented to them.

- The deliverable for this project is a machine learning model and associated web pages for user input and display of results obtained by the model.

B.2 Scope of Project

B.2.A Included in Project Scope

The scope of the project is the creation of a machine learning model (with necessary cleaning, examination, and pre-processing of the dataset) to provide users recommendations of anime based on their preferences, available to them in an interactive webpage.

B.2.B Not included in Project Scope

Not included in this project is a way/method to update the list of available anime, which become available on an ongoing basis. This project will use a fixed collection based on the Kaggle dataset available at the time of download. Additionally, the scope of work does not include a method to interact with a database. Ideally a method to add both new users of the site and their ratings of watched anime would be incorporated into the overall design of the recommender system being developed.

B.3 Standard Methodology

Project Development Methodology: The Agile Model

Generally speaking, Agile is an iterative and incremental approach to software development which emphasizes flexibility, collaboration, and customer feedback. In this project the Agile method is modified to fit a single developer who is also one of the end users of the

project. This iterative approach is suitable for the project as several parts, using divers technologies, must work together for the end result to work properly.

Phases of the Agile Model and their Application to the Project

1) Product Backlog Creation: Formulate a detailed list that captures all the essential features, enhancements, and project requirements. This list should be prioritized based on their criticality, interdependencies, and their potential value to the user. This encompasses all the elements outlined in B.1 and the specifications detailed in B.2.

2) Planning: Strategically dissect the product backlog into smaller, more actionable segments. The inaugural sprint should primarily focus on laying down the foundational infrastructure for the machine learning model and crafting an intuitive web page for user input.

3) Sprint Execution: Dedicate efforts to the specified tasks, which involve rigorous data cleaning, meticulous preprocessing, the initial stages of model training, and the comprehensive design of the web page's user interface.

4) Testing and Feedback: Upon culmination of each sprint, undertake a thorough evaluation of the features that have been developed. In the context of this project, this involves a critical assessment of the operational efficacy of the recommendation system and an examination of the user experience offered by the web page.

5) Iterate: Subject the model and web pages to a stringent functionality test. Implement essential modifications or rectifications based on findings, subsequently updating the product backlog. Strategically plan for the subsequent sprint and reinitiate the iterative process.

6) Release: Upon meticulous testing and validation of a robust set of features, roll them out to the end-users. For this project, this translates into the live deployment of the recommendation system on the website, enabling users to seamlessly engage with it

7) Maintenance and Continuous Improvement: Adopt a proactive approach to constantly monitor the system's performance metrics and actively solicit user feedback. Assimilate this feedback into upcoming sprints with an aim to perpetually refine the system and introduce novel features.

B.4 Timeline and Milestones

Milestone or deliverable	Duration	Projected start	Anticipated end
Deliverable: Interactive Recommender System	15 days	*10/10/2023*	*10/25/2023*
Milestone 1: Data Exploration	4 days	*10/10/2023*	*10/13/2023*
Milestone 2: Model Creation	8 days	*10/14/2023*	*10/21/2023*
Milestone 3: Webpage Creation	1 day	*10/22/2023*	*10/23/2023*
Milestone 4: Test and Deployment	2 days	*10/23/2023*	*10/25/2023*

B.5 Resources and Costs

1. Employee cost: $7,800.00 (15 days of development at $65/hr)
2. Computer Equipment: No cost, already purchased
3. Computational costs: $500.00 (processing time, electrical and cooling, etc.)

Total estimated cost of project: $8,300.00

In the future, similar projects would benefit from upgrading the computer equipment with more powerful GPUs and additional RAM (memory).

B.6 Criteria for Success

The criteria for success of this project is as follows:

1) Working NCF model using the cleaned and processed Kaggle dataset.
2) Webpage to allow user to rate 10 anime (out of the top 30 from the model) on a scale of 1 – 10 (plus 'Haven't seen it") and submit these scores to the model.
3) Webpage that produces a list of 5 anime (as predicted by the model) that the user might enjoy based on their earlier selections.

C. Design of Data Analytics Solution

C.1 Hypothesis

Otaku Castle aims to boost website interaction by addressing the challenge viewers face in discovering new anime. By introducing an anime recommender system using supervised machine learning, where visitors rate a few anime they have previously watched, I hypothesize that the tailored recommendations will not only enhance user engagement but also position Otaku Castle as a "go-to" site for anime enthusiasts seeking personalized content suggestions.

C.2 and C.2.A Analytical Method

My choice of using Neural Collaborative Filtering (NCF) to implement an anime recommender system supports the hypothesis by providing a way to enhance user engagement by not only increasing user interaction but by personalizing the users experience.

Using Neural Collaborative Filtering (NCF) is an appropriate choice because it provides a more robust infrstructure than other solutions. Namely, an NCF model can capture more

complex patterns in user-item interactions. It can scale without significant degradation in performance as Otaku Castle grows and incorporates new data into the model. Further, NCF can be integrated with other neural network architectures or add additional input features if needed in the future. For example, as the store portion of the site grows, purchasing data can be integrated and used to enhance predictions.

Overall, using an NCF model not only provides the granularity of recommendations that are personalized for the users, but also ensures scalability, flexibility, and robustness as Otaku Castle grows and evolves.

C.3 Tools and Environments

The tools used to create the solutions in this project are as follows:

1) Download of the original dataset from Kaggle: myanimelist-dataset
2) Jupyter Notebooks run within Anaconda to perform web scraping tasks and data exploration, cleaning and pre-processing:
    a) Pandas
    b) Numpy
    c) Matplotlib
    d) Beautifulsoup4
    e) Scikit-Learn
    f) PyTorch (with CUDA support)
3) Microsoft Visual Studio 2022 to write the necessary Python (version 3.9) scripts:
    a) PyTorch (with CUDA support) to create and run the NCF model.
    b) Flask to process and display the webpages that provide the model with user ratings, supply those ratings to the model, and provide recommendations to the user after the model has made its predictions.

C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance

Model Description and Metrics

- **The type of model:** Supervised Learning - Regression: The model predicts continuous rating values based on user-anime interactions.
- **The algorithm:** Neural Collaborative Filtering (NCF)
- **The metrics:** Loss, which measures the disparity between predicted and actual user ratings.
- **The benchmark:** Aim for a training loss near 1.6 and ensure the validation loss stays close to training loss values.

Justification of Methods and Metrics

The task at hand is to formulate a recommendation system that provides users with anime suggestions based on their preferences. This task necessitates the prediction of continuous values, i.e., ratings, positioning it within the realm of regression.

Choice of Model: Neural Collaborative Filtering (NCF)

NCF is especially tailored for recommendation systems, seamlessly blending matrix factorization and deep learning techniques. This combination offers the dual advantage of capturing explicit patterns from user-item matrices and subtle nuances through deep learning. The large volume of anime available for viewing necessitates a model that can discern intricate patterns and offer personalized recommendations. NCF's inherent design is aptly equipped for such tasks, making it the best choice.

Metric: Loss

In the domain of machine learning, particularly when predictions involve continuous values, loss provides an invaluable measure of a model's accuracy. By quantifying the difference between predicted and actual values, loss offers a clear performance metric that can be continually optimized. Given the project's aim to provide precise anime recommendations, it's paramount that the model's predictions closely align with actual user ratings. Thus, a metric like loss, which directly reflects this alignment, is both appropriate and indispensable.

Benchmark for Success:

The benchmark values set for training and validation loss are grounded in the objective to ensure both precision in predictions and generalizability to new data. By targeting a training loss near 1.6 and closely monitoring the validation loss, we ensure that the model not only learns effectively from the provided data but also remains adept at handling novel, unseen user-anime interactions.

The choices made in terms of model, metrics, and benchmarks resonate with the overarching objective of the project: to offer users of Otaku Castle personalized and precise anime recommendations. The methods and metrics chosen not only align with the technical requirements of the task but also ensure that the end-users receive recommendations that truly mirror their preferences, thereby enhancing their experience on the platform.

C.5 Practical Significance

In the sprawling universe of anime, potential viewers often grapple with the sheer abundance of choices. Otaku Castle's vision of heightening website interactions is materialized through the inception of an anime recommender system. Harnessing the power of supervised machine learning, the system illuminates a path for visitors, guiding them through a curated journey based on their unique tastes and preferences.

Criteria for Assessing Practical Significance:

- User Engagement: A tangible uptick in user interaction on the website post-implementation of the recommender system. Enhanced user engagement is an indicator of the system's efficacy in tailoring recommendations that resonate with individual preferences.
- User Retention: Monitoring the return rate of users can offer insights into the recommender system's success. A higher retention rate would be indicative of users finding value in the personalized recommendations, fostering loyalty to Otaku Castle.
- Feedback and Reviews: User testimonials and feedback can serve as qualitative measures of the system's impact. Positive reviews centered around the accuracy and relevance of the recommendations can affirm the practical significance of the model.
- Conversion Rates: Monitoring any potential increase in purchases or other desired user actions directly resulting from the recommendations. This would be a direct measure of the system's impact on user behavior.

Much of the criteria will measured through website analytics, and appropriate actions can be taken to improve user satisfaction.

Otaku Castle's aspiration to bolster website interaction is linked to its desire to provide users with a seamless and tailored experience. The recommender system, based on Neural Collaborative Filtering, stands as a testament to this vision. By allowing users to rate a select set of anime, the system, in turn, curates a list of recommendations that echo their unique tastes. The overarching research question centers around the potential of machine learning to enhance user engagement. The implemented system directly addresses this, aiming to transform Otaku Castle into a revered sanctuary for anime aficionados.

Imagine a user, **Cera**, a fan of the fantasy genre. On her first visit to Otaku Castle, she rates a few of her favorite fantasy anime. The recommender system, harnessing the power of NCF, curates a list of lesser-known fantasy anime that aligns with her tastes. Intrigued, **Cera** delves deeper, spending more time on the site. On her subsequent visits, she not only rates more anime but also explores merchandise related to her recommendations. The system's precision in pinpointing her preferences ensures her loyalty to Otaku Castle, and she soon becomes a regular visitor **and buyer**, trusting the platform for her anime recommendations **and merchandise**.

C.6 Visual Communication

Visualizations of the data are produced by the file ‘AnimeRecommender.ipynb’ and will be provided in Task 3 through screenshots of each visualization. Additionally, a screenshot of the model completing its processing will also be inserted into the Task 3 document. Finally, screenshots of the 2 significant web pages will be taken. These pages are ‘ratings.html’ and watchnext.html.’ The first is where a new user inputs their ratings and submits the form to the model. The second page displays the output of the model’s predictions. The screenshots will be

processed using MS Paint, and saved as PNG files. These images will be inserted into the Task 3 document.

D. Description of Dataset

D.1 Source of Data

Kaggle: myanimelist-dataset, 3 CSV files - anime-dataset-2023.csv, users-details-2023.csv, and users-score-2023.csv.

D.2 Appropriateness of Dataset

This data is appropriate for supporting the organizational need because it is explicitly a collection of anime titles and the ratings each have received from users. While there is much unnecessary data within these 3 datasets, at least for the goals of this project, the CSV files are the best collection of anime related information available for public use.

D.3 Data Collection Methods

The data was collected by downloading the three .csv files from Kaggle at: https://www.kaggle.com/datasets/dbdmobile/myanimelist-dataset

D.4 Observations on Quality and Completeness of Data

The amount of data within the three datasets went far beyond the scope of this project. The datasets themselves required little cleaning, mainly centered around ‘UNKNOWN’ values that had to be a number for the model. There was also a significant number of columns between

the three datasets that, while useful information, did not provide information needed for the recommendation system.

It was discovered during the cleaning and exploration phase that users-details-2023.csv was unnecessary. That file contained information that would have been useful had the project wanted to provide some sort of recommendations based on the country, age, or gender of the user, but there was a significant amount of missing data within these columns. However, during cleaning, I was able to confirm that all the users in both the ‘scores’ and ‘details’ datasets matched. This made eliminating the ‘details’ dataset easy as there was no need to combine any of the data into the ’scores’ dataset.

The ‘scores’ dataset was clean, but included columns not necessary for the end result of the project. In order to reduce the file size (and increase processing speed at the same time), these columns were dropped and a new CSV was created for use when running the model.

The ‘anime’ dataset includes a column called ‘Image URL’ which provides a link to an image for each anime in the dataset. This column pointed to the original location of the images, hosted on the My Anime List server. In order to make the images available on the Otaku Castle website (eliminating a lookup on a website outside the control of Otaku Castle) a webscraping script (Pull_Images.ipynb) was created and executed. The script is only run once, at the beginning. This script:

1) Creates an directory to store the images.
2) Downloads each image (at 3 second intervals).
3) Renames the image to the ‘English name’ title of the anime.
4) Updates the ‘Image URL’ column with new site and image name information.
5) Saves the updated dataset to a new CSV for use with the remainder of the project.

After this step is complete, the remainder of the cleaning, exploration and visualization is performed within AnimeRecommender.ipynb and a new CSV is written for use by the model.

D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances

Regarding the concerns in this section, at this stage there are no real issues. The data comes from Kaggle, and is free to use. The only precaution necessary is to remove any Personally Identifiable Information (PII) from the original CSV files. However, in the future, as the project is fully implemented on the Otaku Castle website, these concerns become valid.

Here is a summary of each concern and how it relates to this project (as the project stands upon submission):

- Data governance: Since the dataset is from Kaggle and is public domain, there's no concern about its origin or trustworthiness.
- Privacy: The project only deals with historical data and does not save any Personal Identifiable Information (PII) of new users. This minimizes privacy concerns.
- Security: Again, the project is based on publically available data. At this time, there are no security issues.
- Ethical, legal, and regulatory compliance considerations: Even though the dataset is public domain, it's still crucial to use the data responsibly. The original data included Username, Location, Gender, etc. This data has been scrubbed from the CSV files created to run the model and perform the predictions during the exploration and cleaning phase.

As stated above, it is necessary to remove any potential PII from the data sets, the original CSV files from Kaggle. In this implementation, much of the user information that could be used to find PII is stored within the users-details-2023.csv file. During the exploration and cleaning phase of this project it was determined that the entire CSV file was not relevant to the creation of the model or in providing recommended anime to the new users. As such, this file (and its associated dataframe) is not used past the cleaning phase. The user-scores-2023.csv file only included a Username, which was subsequently dropped as not useful to the model during the creation of the cleaned CSV files.

In the future, Otaku Castle plans to enable users to create accounts. This would facilitate the storage of ratings, plus enable recommendations for products within the store to users based on their personal ratings of anime they have watched, along with several other features not currently available. It is at that point all the above concerns of Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances come into consideration.

As the project evolves and potentially integrates with other datasets or collects new data, Data Governance will become an issue. It will be essential to establish clear, documented rules and guidelines on data collection, storage, and usage. Procedures for data quality checks, data integration, and updates would be outlined and implemented.

Privacy and Security requirements will change as well, and Otaku Castle will implement strong encryption methods for data at rest and in transit. Regular security audits and vulnerability assessments will be conducted, along with ensuring that best practices for password storage is followed both on internal databases and user accounts.

Ethical considerations as Otaku Castle grows revolve around ensuring that data and analytics are not misused to harm individuals or groups. This is not really a concern for a (currently) small company. Misuse of customer data would not only be unethical, but would destroy the business.

To comply with Legal and Regulatory issues, a mechanism to handle GDPR and CCPA regulations will have to be implemented. Further, the Privacy Policy will have to be updated to include what type of data is collected, when, and why. This provides transparency in data collection and usage. Users should know how their data is being used and for what purpose, which in turn, increases trust, heightens user retention, and most importantly, is a driver for increased revenue.

References

Arasanipalai, A. U. (2019, February 14). *How companies use collaborative filtering to learn exactly what you want*. Retrieved from FreeCodeCamp.org: https://www.freecodecamp.org/news/how-companies-use-collaborative-filtering-to-learn-exactly-what-you-want-a3fc58e22ad9/

Nvidia. (2023, 09 30). *Recommendation System*. Retrieved from Nvidia Web Site: https://www.nvidia.com/en-us/glossary/data-science/recommendation-system/

Sharma, A. (2019, December 16). *Neural Collaborative Filtering Supercharging collaborative filtering with neural networks*. Retrieved from TowardsDataScience.com: https://towardsdatascience.com/neural-collaborative-filtering-96cef1009401