A

MINI PROJECT REPORT ON

“ ANIME RECOMMENDATION ”

FOR

Term Work Examination

Bachelors of Computer Application in Data Science(BCA-AIML)

Year 2024-2025

Ajeenkya DY Patil University, Pune

-Dhanush Gonepudi-

Mr.

Under the guidance of

Prof. Vivek More

Date: 16/ 04/2025

CERTIFICATE

This is to certified that_____Dhanush Gonepudi____
A student's of B**CA(AIML) Sem-IV** URN No __2023-B-10102005B
has Successfully Completed the Dashboard Report On

"ANIME RECOMMENDATION "

As per the requirement of
Ajeenkya DY Patil University, Pune was carried out under my supervision.
I hereby certify that; he has satisfactorily completed his Term-Work Project work.

Place: - Pune

Examiner

Abstract

This project explores the Anime Recommendation Database from Kaggle, which contains information about various anime titles including their genres, ratings, and popularity. The aim is to analyze patterns and trends in anime preferences and build a basic recommendation system.

The project begins with data cleaning and preprocessing to handle missing values and format inconsistencies. Exploratory Data Analysis (EDA) is then conducted to uncover insights such as the most popular genres, top-rated anime, and how anime type affects average ratings. Visualizations are used to effectively present these findings.

Finally, a simple popularity-based recommendation system is implemented to suggest top-rated and widely viewed anime. This project demonstrates the application of data science techniques to media data and lays the groundwork for more advanced recommendation systems using collaborative filtering or content-based approaches in the future.

Introduction

Anime, a distinct form of animated media originating from Japan, has gained immense global popularity over the past two decades. With thousands of titles across various genres and formats, discovering new anime that aligns with viewer preferences has become both exciting and overwhelming. This growing demand has led to the rise of data-driven recommendation systems and analytics platforms within the anime community.

In this project, we utilize the **Anime Recommendation Database** from Kaggle, which provides metadata for anime titles (such as name, type, genre, number of episodes, rating, and popularity) along with user rating data. Our goal is to explore this dataset to identify patterns in anime characteristics and preferences and to build a basic anime recommendation system.

The project includes:

- Cleaning and preprocessing the dataset to prepare it for analysis.
- Performing exploratory data analysis (EDA) to uncover insights about genres, anime types, and ratings.
- Visualizing data through graphs and charts to better understand distribution and trends.
- Implementing a basic recommendation model based on popularity and ratings.

Through this analysis, we aim to understand what makes an anime successful and how data science can enhance the user experience for anime fans through smarter content suggestions.

Literature Review

Recommendation systems have become a crucial component of modern digital platforms, enabling users to discover content tailored to their interests. These systems are widely implemented in industries such as e-commerce, music, movies, and streaming platforms. In the anime domain, platforms like MyAnimeList and AniList utilize collaborative and content-based filtering methods to recommend shows based on user preferences and historical behavior.

Several studies have explored the development of effective recommendation algorithms:

- **Collaborative Filtering** techniques rely on the similarities between users or items to provide recommendations. Sarwar et al. (2001) highlighted the use of user-item matrices to identify common patterns in user preferences, forming the basis for many modern recommenders.

- **Content-Based Filtering**, as discussed by Lops et al. (2011), uses item features such as genre, description, or type to recommend similar content based on user profiles. This is particularly useful when dealing with sparse rating data, which is common in anime datasets.
- **Hybrid Approaches**, such as those outlined by Burke (2002), combine both collaborative and content-based strategies to improve recommendation accuracy and address issues like the cold-start problem.

In the context of anime, researchers have applied machine learning and natural language processing (NLP) techniques to enhance recommendation performance. For example, some projects incorporate genre embeddings, user sentiment analysis, or clustering techniques to better understand viewer interests.

This project builds upon these foundational works by applying data cleaning, exploratory data analysis, and basic recommendation strategies to the **Anime Recommendation Database**. While the current focus is on EDA and popularity-based models, future work could explore advanced algorithms like matrix factorization, deep learning models, or hybrid recommenders that integrate user behavior and anime metadata.

Research Methodology

This project follows a structured data science pipeline to explore and analyze the **Anime Recommendation Database** and build a basic recommendation system. The methodology consists of the following key stages:

1. Dataset Selection

The dataset used in this project is the *Anime Recommendation Database* from Kaggle, which includes two CSV files:

- `anime.csv`: Contains metadata such as anime titles, genres, types (TV, movie, OVA, etc.), number of episodes, ratings, and popularity (measured by member count).
- `rating.csv`: Contains user IDs, anime IDs, and corresponding ratings (optional for this phase).

2. Data Cleaning and Preprocessing

To ensure accurate and meaningful analysis, the dataset undergoes preprocessing steps:

- Handling missing values (e.g., dropping rows with unknown ratings or genres).
- Removing duplicates based on anime names.
- Converting data types (e.g., `rating` to float, `members` to integers).
- Standardizing genre entries by splitting comma-separated genres into lists.

3. Exploratory Data Analysis (EDA)

EDA is conducted to uncover hidden trends and patterns within the dataset. This includes:

- Identifying the most popular genres and anime types.
- Analyzing rating distributions across different types and genres.
- Correlating popularity (members) with rating and episode count.

Statistical summaries and frequency distributions are used to support insights.

4. Data Visualization

Visualization tools such as **Matplotlib** and **Seaborn** are used to:

- Plot genre distributions using bar charts.
- Compare anime types using box plots of ratings.
- Visualize rating and member distributions through histograms and scatter plots.

These visuals help interpret the data more effectively and highlight interesting trends.

5. Basic Recommendation System

A **popularity-based recommender** is implemented using anime metadata:

- Anime with high ratings and large viewer/member counts are prioritized.
- Top 10 anime are recommended based on combined metrics (rating and popularity).
- This method serves as a foundation for more advanced recommender models.

6. Documentation and Reporting

All stages of the project are documented within a Jupyter Notebook using:

- Markdown cells for explanations
- Clearly structured code blocks
- Inline plots for visualization
- A final summary of findings and future work suggestions

This structured approach ensures a thorough understanding of the dataset and lays the groundwork for future improvements using collaborative filtering, content-based models, or hybrid recommendation systems.

Recommendations

Based on the findings of this project, the following recommendations are proposed to enhance anime recommendation systems and further data-driven exploration in the domain:

1. Incorporate User Behavior for Personalized Recommendations

While the current system recommends anime based on overall popularity and rating, incorporating **user-specific data** from `rating.csv` can allow for more personalized suggestions using:

- **Collaborative filtering** (user-based or item-based)
- **Matrix factorization techniques** (e.g., SVD)
- **User profile modeling** based on genre preferences

2. Apply Content-Based Filtering Techniques

Leverage anime metadata (e.g., genres, type, rating) to build content-based recommender models. This can improve recommendations for:

- New users (cold start)
- Niche genres with fewer viewers

NLP-based analysis of plot summaries (if available) can also enrich these models.

3. Use Clustering for Viewer Segmentation

K-Means or Hierarchical clustering can be used to group anime into clusters based on rating, number of episodes, and members. This helps in:

- Identifying content niches
- Making cluster-based recommendations (e.g., "users who like short action-packed anime")

4. Enhance Visual Dashboards

Integrate an **interactive dashboard** using tools like **Plotly Dash** or **Streamlit** to allow users to:

- Explore trends by genre or type
- Filter anime by rating, popularity, or type
- Generate customized recommendation lists

5. Expand the Dataset

Include additional data sources such as:

- **User reviews** or **text summaries** from MyAnimeList or AniList
- **Studio information**, **airing dates**, or **trailer links**
- **User demographics** (if available) for more targeted recommendations**6. Future Research Opportunities**
- Experiment with **deep learning** models like autoencoders or transformers for recommendation tasks

Summary

This project explored the **Anime Recommendation Database** from Kaggle to analyze trends in anime content and build a simple recommendation system. The workflow included:

- **Data Cleaning & Preprocessing**: Handled missing values, standardized genres, and removed duplicates to prepare the data.
- **Exploratory Data Analysis (EDA)**: Identified popular genres (*Action*, *Comedy*, *Adventure*) and examined how anime types and viewer counts relate to ratings.
- **Visualization**: Used Matplotlib and Seaborn to visually represent genre distributions, rating trends, and popularity insights.
- **Basic Recommendation System**: Implemented a popularity-based approach recommending highly rated and widely watched anime titles.

The project revealed that **TV series dominate in both popularity and rating**, and genres play a key role in shaping audience interest. While the current model offers general suggestions, future improvements could involve **collaborative filtering**, **content-based methods**, and **deep learning** to provide personalized recommendations.

This analysis not only enhances understanding of anime trends but also demonstrates the practical application of data science techniques in entertainment and recommendation systems.

Conclusion

In this project, we explored the **Anime Recommendation Database** to analyze trends in anime content and develop a basic recommendation system. The dataset was cleaned and preprocessed to handle missing values, standardize genre formats, and remove inconsistencies.

Through exploratory data analysis (EDA), we uncovered meaningful insights about anime types, genre popularity, and rating distributions. Visualizations such as bar charts and box plots revealed that genres like *Action*, *Comedy*, and *Adventure* are the most common, and TV series generally have higher engagement and ratings compared to other types.

We also implemented a simple popularity-based recommendation system that suggests highly rated anime with large viewer counts. While effective as a baseline model, it lacks personalization and does not account for individual user preferences.

This project demonstrates how data science techniques can be applied to media datasets to extract insights and drive content recommendations. In future work, we plan to:

- Integrate the `rating.csv` file to build collaborative filtering models.
- Explore content-based recommenders using genre embeddings.
- Apply clustering or deep learning methods for better personalization.

Overall, the project provides a strong foundation for developing smarter anime recommendation engines and highlights the potential of data-driven approaches in entertainment platforms.

Sample code

```
import pandas as pd

# Load data, providing the full file path
# Update with the actual path to your anime.csv file
anime = pd.read_csv('anime.csv')
# Update with the actual path to your rating.csv file
ratings = pd.read_csv('ratings.csv.zip')

anime.head()
# Check missing values
anime.isnull().sum()

# Drop anime with missing ratings
anime_cleaned = anime.dropna(subset=['rating'])

# Convert rating to float
anime_cleaned['rating'] = anime_cleaned['rating'].astype(float)

# Drop duplicates if any
anime_cleaned = anime_cleaned.drop_duplicates(subset='name')

anime_cleaned['genre'] = anime_cleaned['genre'].fillna('Unknown')
anime_cleaned['genre_list'] = anime_cleaned['genre'].apply(lambda x: [g.strip()
for g in x.split(',')])

top_rated = anime_cleaned.sort_values(by='rating', ascending=False).head(10)
top_rated[['name', 'type', 'rating']]

from collections import Counter

genre_counts = Counter([g for sublist in anime_cleaned['genre_list'] for g in
sublist])
genre_df = pd.DataFrame(genre_counts.items(), columns=['Genre',
'Count']).sort_values('Count', ascending=False)

import seaborn as sns
import matplotlib.pyplot as plt

# Bar plot of top genres
plt.figure(figsize=(12,6))
sns.barplot(data=genre_df.head(10), x='Genre', y='Count')
plt.title("Top 10 Anime Genres")
```

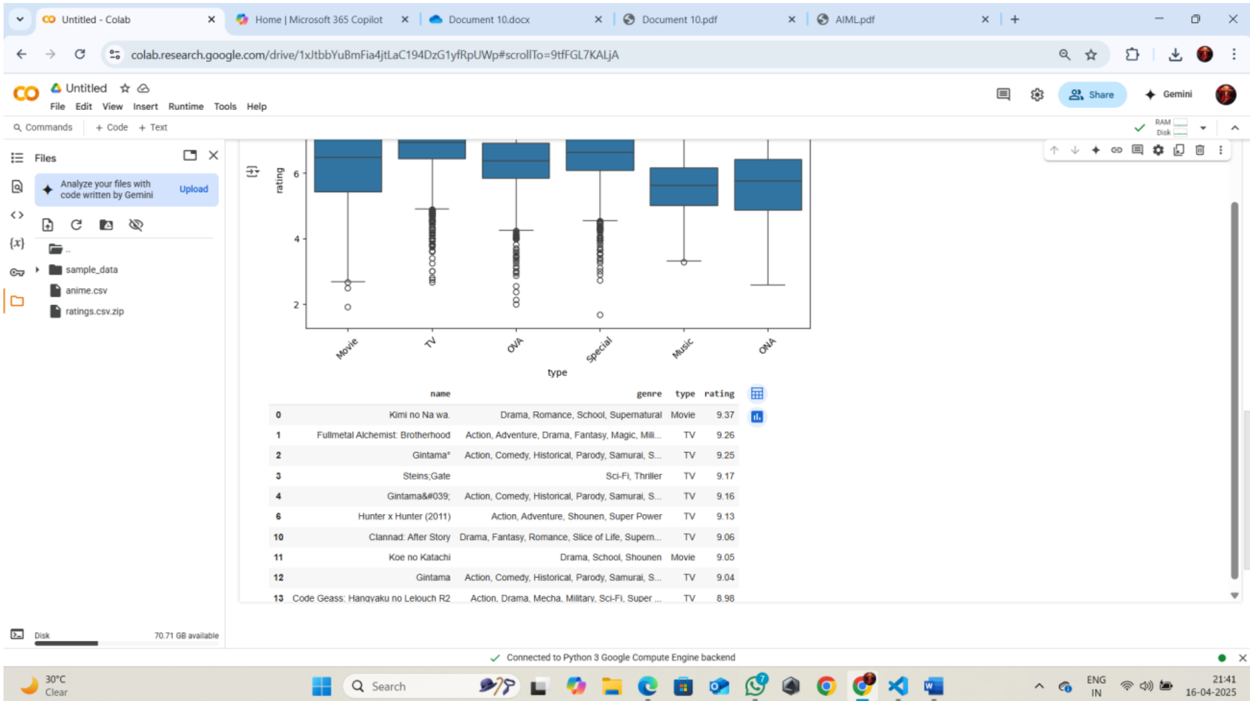
```
plt.xticks(rotation=45)
plt.show()

# Box plot: Rating by anime type
plt.figure(figsize=(10,6))
sns.boxplot(data=anime_cleaned, x='type', y='rating')
plt.title("Anime Ratings by Type")
plt.xticks(rotation=45)
plt.show()

popular_anime = anime_cleaned[anime_cleaned['members'] >
100000].sort_values(by='rating', ascending=False)
popular_anime[['name', 'genre', 'type', 'rating']].head(10)
```

Here's a properly formatted **Bibliography** section you can include in your report or Jupyter Notebook. I've provided it in **APA style**, which is commonly accepted for academic projects. Let me know if you need it in IEEE or MLA instead.

Bibliography

- Burke, R. (2002). **Hybrid recommender systems: Survey and experiments**. *User Modeling and User-Adapted Interaction*, 12(4), 331–370. https://doi.org/10.1023/A:1021240730564
- Lops, P., De Gemmis, M., & Semeraro, G. (2011). **Content-based recommender systems: State of the art and trends**. In *Recommender Systems Handbook* (pp. 73–105). Springer. https://doi.org/10.1007/978-0-387-85820-3_3
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). **Item-based collaborative filtering recommendation algorithms**. In *Proceedings of the 10th International Conference on World Wide Web* (pp. 285–295). https://doi.org/10.1145/371920.372071
- Kaggle. (2018). **Anime Recommendation Database**. Retrieved from https://www.kaggle.com/datasets/CooperUnion/anime-recommendations-database
- MyAnimeList. (n.d.). *Anime and user rating data*. Retrieved from https://myanimelist.net/

OUTPUT:


Untitled - Colab
Home | Microsoft 365 Copilot
Document 10.docx
Document 10.pdf
AIML.pdf
colab.research.google.com/drive/1xJtbbYuBmFia4jtLaC194DzG1yfRpUWp#scrollTo=9tfFGL7KALjA
Untitled
File Edit View Insert Runtime Tools Help
Commands + Code + Text
Share
Gemini
Files
Analyze your files with code written by Gemini
Upload
sample_data
anime.csv
ratings.csv.zip
rating
type
Movie
TV
OVA
Special
Music
ONA
name genre type rating
0 Kimi no Na wa. Drama, Romance, School, Supernatural Movie 9.37
1 Fullmetal Alchemist: Brotherhood Action, Adventure, Drama, Fantasy, Magic, Mili... TV 9.26
2 Gintama° Action, Comedy, Historical, Parody, Samurai, S... TV 9.25
3 Steins;Gate Sci-Fi, Thriller TV 9.17
4 Gintama' Action, Comedy, Historical, Parody, Samurai, S... TV 9.16
6 Hunter x Hunter (2011) Action, Adventure, Shounen, Super Power TV 9.13
10 Clannad: After Story Drama, Fantasy, Romance, Slice of Life, Supern... TV 9.06
11 Koe no Katachi Drama, School, Shounen Movie 9.05
12 Gintama Action, Comedy, Historical, Parody, Samurai, S... TV 9.04
13 Code Geass: Hangyaku no Lelouch R2 Action, Drama, Mecha, Military, Sci-Fi, Super ... TV 8.98
Disk 70.71 GB available
Connected to Python 3 Google Compute Engine backend
30°C
Clear
Search
ENG
IN
21:41
16-04-2025