

# **SEMESTER PROJECT FINAL REPORT**

## **ANIME RECOMMENDATION SYSTEM**

### **Group Members**

Muhammad Ammar Kazmi	70139353@student.uol.edu.pk
Shahzaib Khan	70138852@student.uol.edu.pk
Hafiz Muhammad Tariq	70139976@student.uol.edu.pk

Submitted to:

Sir Asif Ahsan

# 1. Problem Formulation

With the rapid growth of digital entertainment platforms, users are exposed to an overwhelming amount of content. In the anime domain, viewers often face difficulty discovering new anime titles that align with their preferences due to the vast and continuously expanding anime library. Manually searching through thousands of titles is time-consuming and inefficient, leading to information overload and reduced user satisfaction.

The core problem addressed by this project is **effective content discovery**. Specifically, when a user enjoys a particular anime, it becomes challenging to identify other anime with similar characteristics such as genre or thematic elements. As a result, users may miss out on content that closely matches their interests.

---

# 2. Problem Statement

Given a large dataset of anime titles along with associated metadata (name, genre, and description), the objective is to develop a recommendation system that can accurately suggest anime titles similar to a given anime. The system should automatically identify relevant anime and enhance the user's viewing experience by facilitating efficient and personalized content discovery.

---

# 3. Existing Solutions

Existing recommendation approaches for content discovery mainly include **content-based filtering**, **collaborative filtering**, and **hybrid systems**. **Content-based filtering** suggests items similar to those a user already likes by analyzing item attributes such as genre, name, or textual features; for example, if a user enjoys action fantasy anime, the system recommends titles with similar characteristics, making this approach well-suited for situations where detailed metadata is available. **Collaborative filtering**, on the other hand, relies on user interaction data such as ratings or viewing history to identify patterns among users or items, but it becomes ineffective when such data is unavailable or sparse. Hybrid recommendation systems combine both content-based and collaborative techniques to overcome individual limitations, particularly the cold-start problem; however, due to the lack of user behavior data, this project primarily focuses on a **content-based filtering** approach using **TF-IDF** and **cosine similarity** to generate relevant anime recommendations.

---

## 4. Proposed Solution

The proposed solution is a content-based anime recommendation system that identifies and suggests anime titles similar to a given input anime by analyzing their textual features. The system combines important metadata such as anime name and genre into a unified textual representation and applies TF-IDF vectorization to convert this information into numerical feature vectors. Cosine similarity is then used to measure the degree of similarity between anime titles based on these vectors. For a selected anime, the system retrieves and ranks the most similar titles, ensuring relevant and meaningful recommendations. This approach does not rely on user interaction data, making it suitable for scenarios with limited or no user history, while still providing accurate and efficient recommendations. The proposed system is simple, scalable, and effective for enhancing content discovery in large anime datasets.

---

## 5. Methodology

### Step 1: Data Loading and Exploration

- **Dataset:** Anime dataset (anime.csv).
  - **Records:** Thousands of anime titles with user ratings.
  - **Attributes:** Anime name, genre, type, episodes, description, and user ratings.
- 

### Step 2: Data Cleaning and Balancing

- Removed missing or duplicate entries.
  - Handled null values in names and genre fields.
  - Ensured fair representation of different anime genres.
- 

### Step 3: Data Preprocessing

- Converted categorical data into machine-readable format.
- 

### Step 4: Feature Extraction

- Transformed anime attributes into feature vectors.
  - Applied techniques such as:
    - TF-IDF for genre and name
- 

### Step 5: Model Training

- **Feature Combination:**  
Combining relevant text fields like name and genre into a single string
  - **TF-IDF Vectorization:**
-

Using the Tf-idf Vectorizer to convert the combined text initial into numerical vectors.

---

### **Step 6: Model Evaluation**

- Evaluated system performance using:
  - Precision of suggested anime
  - Achieved high recommendation relevance and consistency
- 

### **Step 7: Model Deployment and Testing**

- Saved trained recommendation models.
  - System correctly generated personalized anime recommendations.
- 

## **6. Implementation and Experiments**

### **6.1 Implementation Details**

- Key libraries include **Pandas**, **NumPy**, and **scikit-learn**.
  - TF-IDF vectorization and cosine similarity are implemented for feature extraction and similarity computation.
- 

### **6.2 Experimental Setup**

#### **Experiment 1: Initial Model**

- Features: Anime name and genre
- TF-IDF limit: `max_features = 2000`
- Relevance criterion: At least two common genres
- **Result:**  
`Precision@5 = 0.50`

#### **Experiment 2: Improved Model**

- TF-IDF without feature limitation
  - Relevance criterion relaxed to at least one common genre
  - **Result:**  
`Precision@5 = 0.8`
-

## 7. Results

Model Version	Precision@5
Initial Model	0.50
Improved Model	0.80

The improved model provides more contextually relevant anime recommendations and better aligns with user expectations.

---

## 8. Conclusion

This project successfully implemented a **content-based anime recommendation system** using TF-IDF vectorization and cosine similarity. The system effectively addresses the challenge of content discovery in large anime datasets.

---

### Key Findings

- Content-based filtering is effective when quality metadata is available.
  - The system is simple, interpretable, and easy to implement.
  - Pre-computed similarity matrices allow fast recommendation retrieval.
- 

### Limitations

- Overspecialization due to reliance on content similarity
  - Lack of personalization due to absence of user behavior data
- 

### Future Work

- Incorporating anime descriptions, tags, and studio information
  - Exploring alternative similarity measures
  - Implementing hybrid recommendation techniques for improved personalization
-

## 9. References

- [1] M. K. Sidhu and N. Sharma, “Enhancing personalized movie recommendations with cosine similarity: A genre and feature-based approach,” *International Journal of Engineering Research & Technology (IJERT)*, vol. 14, no. 2, Feb. 2025.
- [2] R. Huang, “Improved content recommendation algorithm integrating semantic information,” *Journal of Big Data*, vol. 10, art. 84, May 2023.
- [3] F. Christyawati, A. N. Rohman, and A. D. Hartanto, “Application of content-based filtering method using cosine similarity in restaurant selection recommendation system,” *Journal of Information Systems and Informatics*, vol. 6, no. 3, pp. 1559–1576, 2024.