



Project Report

Netflix Movie Recommendation System

1. Introduction

Online streaming platforms such as Netflix host thousands of movies and TV shows. With such a large content library, helping users find relevant and engaging content becomes essential. Recommendation systems address this need by automatically suggesting movies or shows that match user interests.

This project develops a **Content-Based Movie Recommendation System** that recommends movies similar to a user-liked movie. The system analyzes movie features such as genre, cast, director, and description, and identifies similar content using Natural Language Processing (NLP) techniques and similarity algorithms.

2. Problem Statement

Users often struggle to decide what to watch next due to the massive availability of video content. The problem is to design a system that:

- Understands what a movie is about
 - Finds other movies with similar characteristics
 - Provides accurate and meaningful recommendations
-

3. Objectives

The main objectives of this project are:

1. To understand and implement Content-Based Recommendation Systems
 2. To preprocess and clean textual movie metadata
 3. To convert text information into numerical features using NLP
 4. To compute similarity between movies
 5. To recommend Top-N similar movies based on user input
 6. To evaluate recommendation performance
-

4. Dataset Description

The dataset used is the **Netflix Movies and TV Shows Dataset** from Kaggle.

Dataset Source:

Kaggle – Netflix Movies and TV Shows Dataset

Contains details of all movies and shows available on Netflix.

Dataset Attributes

- show_id – Unique ID
 - title – Movie/Show Name
 - director – Director Name
 - cast – Actors/Actresses
 - country – Country Produced
 - date_added – Added on Netflix
 - release_year – Release Year
 - rating – Age Rating
 - duration – Movie length / episodes
 - listed_in – Genre / Category
 - description – Short Summary
-

5. Methodology / Project Workflow

Step 1: Data Acquisition

The dataset was downloaded from Kaggle and loaded using Pandas.

Step 2: Data Preprocessing

Performed:

- Handling missing values
- Selecting useful features
- Creating a combined feature “content”

Selected columns:

- title
- director
- cast
- listed_in
- description

Missing values were replaced with blank spaces to avoid data loss.

Step 3: Feature Engineering

A new feature called **content** was created by combining:

director + cast + genre + description

This provides a rich textual profile for every movie or show.

Step 4: Text Vectorization

Movies' textual data was converted into numerical form using **TF-IDF Vectorization**.

Why TF-IDF?

- Assigns more weight to meaningful words
 - Reduces impact of common words
 - Improves recommendation accuracy
- Compared to CountVectorizer, TF-IDF provides better feature representation.
-

Step 5: Similarity Computation

Similarity between movies was calculated using **Cosine Similarity**.

Why Cosine Similarity?

- Measures similarity between text vectors
- Works best for NLP and recommendation systems
- Ignores difference in document length

This produces a similarity score matrix where:

`similarity[i][j]`

= similarity between movie i and movie j

Step 6: Recommendation Function

A function was built that:

- Takes a movie name as input
- Finds its index
- Fetches similarity scores
- Sorts movies by similarity

- Returns Top-N similar movies
-

6. Technologies Used

- Python
 - Pandas
 - NumPy
 - Scikit-Learn
 - NLP Techniques
 - TF-IDF Vectorizer
 - Cosine Similarity
-

7. Results

The recommendation system successfully provides meaningful movie recommendations based on user input.

Example Output

Input:

"Money Heist"

Output Recommendations:

- Narcos
- Breaking Bad
- Queen of the South
- Ozark
- Peaky Blinders
- El Chapo
- Prison Break
- Blacklist
- Shooter
- Power

The results show that the system correctly identifies crime-thriller and drama-based shows similar to Money Heist, proving good recommendation quality.

8. Evaluation

Evaluation was performed through:

- Manual inspection
- Logical similarity validation
- Feature comparison tests

Experiments were done using:

- Only description
- Combined features
- TF-IDF vs CountVectorizer

Combined features + TF-IDF showed the best results.

9. Applications

This system can be used in:

- OTT platforms like Netflix, Amazon Prime, Hotstar
 - Movie streaming applications
 - Personalized content platforms
 - Entertainment portals
-

10. Conclusion

This project successfully implements a **Content-Based Movie Recommendation System** for Netflix Movies and TV Shows. The system analyzes textual metadata, converts it into numerical features, and recommends similar movies using cosine similarity.

The project enhances understanding of:

- Recommendation System Concepts
 - Natural Language Processing
 - Feature Engineering
 - Similarity Algorithms
 - Practical use of Machine Learning libraries
-

11. Future Enhancements

Possible improvements:

- Hybrid Recommendation System (Content + Collaborative)
 - Sentiment Analysis from user reviews
 - Real-time recommendation updates
 - Web/Android Deployment using Flask / Streamlit
 - Personalized user-based recommendations
-

12. References

1. Kaggle Netflix Dataset
2. Scikit-Learn Documentation
3. NLP & TF-IDF Concepts
4. Cosine Similarity Research Papers