Internship Project Report

Name: Nandini Gupta

Project Title: Movie Recommendation System

Date: July 2025

1. INTRODUCTION

In today's digital world, where hundreds of movies release every year, users often find it difficult to choose what to watch next. A movie recommendation system helps solve this by suggesting movies that match a user's preferences. These systems are widely used on platforms like Netflix, Amazon Prime, and YouTube. The goal of this project is to build a basic yet effective recommendation system using real-world data.

2. ABSTRACT

This project focuses on building a machine learning-based movie recommendation system using the MovieLens dataset. The system analyzes movie genres and user ratings to suggest similar movies. We implemented a content-based filtering approach that compares genre similarities between movies. Python and its data science libraries were used for implementation. The final system takes a movie name as input and returns a list of five similar movies. This project demonstrates the basic working of a recommender engine and how it can be expanded further using collaborative filtering and user behavior.

3. TOOLS USED

- Programming Language: Python
- Libraries: Pandas, Scikit-learn, TfidfVectorizer, Cosine Similarity
- Platform: Google Colab
- Dataset: MovieLens (ml-latest-small)

4. STEPS INVOLVED IN BUILDING THE PROJECT

1. Dataset Collection
   - Downloaded MovieLens dataset containing movie titles, genres, and user ratings.

2. Data Preprocessing
   - Cleaned data by handling missing values and formatting genre text.

3. Content-Based Filtering

- Used TF-IDF (Term Frequency–Inverse Document Frequency) to convert movie genres into numerical vectors.
- Calculated similarity scores between all movies using cosine similarity.

4. Recommendation Logic
- Built a function to find the most similar movies based on genre similarity.
- Returned top 5 movie recommendations for a given input movie title.

5. Testing & Output
- Tested the system with various movie titles like "Toy Story" and "Iron Man" to verify the recommendations.

5. CONCLUSION

The project successfully demonstrates how a simple content-based movie recommendation system can be built using Python and real-world data.The core logic was first developed and tested in Google Colab using a Jupyter notebook. Later, the system was integrated into a Streamlit-based user interface for easier interaction.While this version focuses on genre similarity, future enhancements can include collaborative filtering, sentiment analysis from reviews, and a user interface for better interaction. This project was helpful in understanding how real-world recommendation engines work and how machine learning can be applied to personalized suggestions.