SAVITRIBAI PHULE PUNE UNIVERSITY

ACTIVITY BOOK FOR

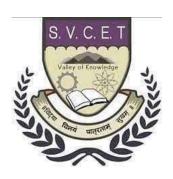
"Content-Based Movie Recommender System"

Submitted to the Department of Computer Engineering, SVCET, Rajuri, Pune, in partial fulfillment of the requirements for the

FINAL YEAR OF COMPUTER ENGINEERING



Group Id: 002



DEPARTMENT OF COMPUTER ENGINEERING

SAHYADRI VALLEY COLLEGE OF ENGINEERING & TECHNOLOGY, RAJURI, PUNE (2024-2025)

Title Page

1. Name of Student : MR. Khedkar Laxman Bhimrao

MR. Hase Ashutosh Prakash MR. Langde Ritesh Vitthalrao

2. **Branch** : Computer Engineering

3. **Academic Year** : 2024-25

4. Activity Book : 16/01/2025 – 12/03/2025

5. **Proposed Topic**: Content-Based Movie Recommender System

Day & Date : Thursday 16/01/2025

Day 1

Work Done: Finalized project topic and objectives for movie recommender system.

Tools Used: Web browser, project planning tools.

Learnings: Understanding scope of recommendation systems.

Challenges: Defining clear goals and scope.

Next Steps: Search for appropriate movie rating datasets.

Day & Date : Friday 17/01/2025

■ Day 2

Work Done: Searched and shortlisted potential datasets (e.g., MovieLens).

Tools Used: Web browser, Kaggle.

Learnings: Dataset availability and licensing considerations.

Challenges: Ensuring dataset size and richness.

Next Steps: Download selected dataset.

Day & Date : Wednesday 22/01/2025

Day 3

Work Done: Downloaded MovieLens dataset and explored file structure.

Tools Used: Python (pandas), Jupyter Notebook.

Learnings: Data formats for ratings, movies, tags.

Challenges: Large file handling and initial loading.

Next Steps: Conduct exploratory data analysis.

Day & Date : Thursday 23/01/2025

.... Day 4

Work Done: Performed initial exploratory data analysis: summary statistics, rating distributions.

Tools Used: Python (pandas, matplotlib).

Learnings: User rating patterns and popular movies.

Challenges: Visualizing large datasets effectively.

Next Steps: Deeper analysis of missing values and outliers.

Day & Date : Friday 24/01/2025

EXECUTE: Day 5

Work Done: Conducted deeper EDA: identified missing values, anomalies, genre distributions.

Tools Used: Python (pandas, matplotlib).

Learnings: Data quality issues and feature distribution

Challenges: Handling sparse user-item interactions.

Next Steps: Begin data cleaning and preprocessing.

.....

Day & Date : Wednesday 29/01/2025

Day 6

Work Done: Cleaned dataset: handled missing values, normalized fields.

Tools Used: Python (pandas, numpy).

Learnings: Effective cleaning techniques.

Challenges: Deciding imputation strategies.

Next Steps: Feature engineering for content attributes.

Day & Date: Thursday 30/01/2025



Work Done: Performed feature engineering: extracted genres into binary features.

Tools Used: Python (pandas).

Learnings: Representing categorical attributes.

Challenges: High-dimensional sparse features.

Next Steps: Process text metadata (e.g., tags) using NLP.

.....

Day & Date : Friday 31/01/2025



Work Done: Processed text metadata using TF-IDF vectorization.

Tools Used: Python (scikit-learn, NLTK)

Learnings: Text vectorization and importance weighting.

Challenges: Balancing computational cost.

Next Steps: Implement collaborative filtering models.

Day & Date : Wednesday 05/02/2025

EXECUTE: Day 9

Work Done: Implemented user-based and item-based collaborative filtering prototypes.

Tools Used: Python (surprise library).

Learnings: Similarity metrics and model basics.

Challenges: Cold-start and scale issues.

Next Steps: Download bill.

.....

Day & Date : Thursday 06/02/2025

Day 10

Work Done: Evaluated CF prototypes: computed RMSE on test split, top-N precision/recall.

Tools Used: Python.

Learnings: Interpreting evaluation metrics, recognizing overfitting.

Challenges: Tuning neighborhood size; dataset split strategies.

Next Steps: Implement matrix factorization (SVD) and compare

Day & Date : Friday 07/02/2025

.... Day 11

Work Done: Implemented matrix factorization (SVD) model for ratings prediction.

Tools Used: Python (surprise or implicit libraries).

Learnings: Latent factor modeling, hyperparameter tuning (factors, regularization).

Challenges: Selecting optimal parameters; convergence time.

Next Steps: Compare SVD results vs. baseline CF.

.....

Day & Date : Wednesday 12/02/2025

Day 12

Work Done: Compared SVD performance with neighborhood CF; documented findings.

Tools Used: Python.

Learnings: Trade-offs between model complexity and accuracy.

Challenges: Unexpected segment performance; cold-start still an issue.

Next Steps: Implement content-based recommendation.

Day & Date : Thursday 13/02/2025

.... Day 13

Work Done: Built content-based recommender using metadata features and TF-IDF vectors.ent test emails with bill info using SMTP.

Tools Used: Python (scikit-learn).

Learnings: Combining genre vectors and text vectors for similarity.

Challenges: Feature weighting balance; threshold setting.

Next Steps: Evaluate content-based model (precision/recall).

.....

Day & Date : Friday 14/02/2025

EXECUTE Day 14

Work Done: Evaluated content-based recommendations: measured relevance and diversity.

Tools Used: Python.

Learnings: Content-based strengths (new items) and limitations (over-specialization).

Challenges: Ensuring novelty vs. similarity balance.

Next Steps: Design hybrid recommendation strategy.

Day & Date : Wednesday 19/02/2025

.... Day 15

Work Done: Designed hybrid approach combining CF and content-based scores.

Tools Used: Python

Learnings: Weighting schemes, blending strategies (e.g., linear combination, switching).

Challenges: Determining optimal blending weights.

Next Steps: Implement hybrid model prototype.

.....

Day & Date : Thursday 20/02/2025

Day 16

Work Done: Implemented hybrid recommender prototype (e.g., weighted sum of normalized scores).

Tools Used: Python.

Learnings: Integration details, performance implications.

Challenges: A Efficient computation of combined scores at scale.

Next Steps: Evaluate hybrid model on test data.

Day & Date : Friday 21/02/2025



Work Done: Evaluated hybrid model: measured RMSE (if applicable), top-N precision/recall, user satisfaction proxies.

Tools Used: Python.

Learnings: Hybrid improvements over pure CF or content-based in cold-start and accuracy.

Challenges: Overfitting popular items; tuning to avoid bias.

Next Steps: Plan user interface for demonstration.

.....

Day & Date : Wednesday 26/02/2025

Day 18

Work Done: Sketched UI/UX flow for recommendation demo (web-based).

Tools Used: Design tool or pen & paper, flowcharts.

Learnings: User interaction considerations (input, output display, explanation).

Challenges: Simplifying complexity for clear demo.

Next Steps: Set up development environment for UI.

Day & Date : Thursday 27/02/2025

Day 19

Work Done: Set up Streamlit (or Flask) environment; created basic app skeleton.

Tools Used: Streamlit, Python.

Learnings: Framework basics, routing, UI components.

Challenges: Initial configuration and dependencies.

Next Steps: Integrate recommendation backend into UI.

.....

Day & Date : Friday 28/02/2025

Day 20

Work Done: Integrated recommendation logic into UI; enabled user input and display of recommendations.

Tools Used: Python, Streamlit.

Learnings: Connecting backend functions to UI elements; handling user inputs.

Challenges: Handling latency on large datasets; caching results.

Next Steps: Test UI with sample users and edge cases.

Day & Date : Wednesday 05/03/2025

Day 21

Work Done: Performed user testing of the prototype; collected feedback on usability and relevance.

Tools Used: Streamlit, simple feedback forms/surveys.

Learnings: User preferences for UI layout, explanation clarity.

Challenges: Users encountering unexpected inputs; performance lags.

Next Steps: Debug and optimize UI/backend, add explanation features.

.....

Day & Date : Thursday 06/03/2025

Day 22

Work Done: Refined UI/backend: added caching, improved response times, implemented recommendation explanations.

Tools Used: Python, Streamlit caching, simple NLP for explanations.

Learnings: Importance of explainability; performance tuning techniques.

Challenges: Generating concise, meaningful explanations.

Next Steps: Prepare documentation and reporting.

Day & Date : Friday 07/03/2025

Day 23

Work Done: Drafted project documentation: methodology, code comments, user guide, and report sections.

Tools Used: MS Word/Markdown, code editor.

Learnings: Clear writing of technical methods and results.

Challenges: Ensuring completeness and clarity for varied audiences.

Next Steps: Final review of deliverables and create presentation.

.....

Day & Date : Wednesday 12/03/2025



Work Done: Conducted final review: polished report, finalized code repository, rehearsed presentation/demo.

Tools Used: MS Word/PowerPoint/Streamlit demo.

Learnings: Importance of thorough testing and rehearsal.

Challenges: Last-minute fixes and time management.

Next Steps: Submit project deliverables and present findings.

