



INNOMATICS

RESEARCH LABS

Project Outline Submission

On

Movie Recommender System

SUBMITTED BY

Koukuntla Komal Reddy

Chalumuri Harshitha

Chintada Jagan Mohan Rao

Koppaka Yeswanth

Kalapala Venkata Surya Teja

Asif Ahmad Najar

BATCH

FEB -23

GUIDED BY

Kanav Bansal (Chief Data Scientist)

Vijit Singh

PROJECT OUTLINE SUBMISSION

Project Goal

To build a movie recommender system that can recommend movies to users based on their past ratings and preferences.

Project Objectives

- To clean and pre-process the data.
- To perform exploratory data analysis.
- To choose the best machine learning algorithms for the task.
- To train the machine learning algorithms on the data.
- To make predictions of movies that users are likely to enjoy.
- To evaluate the performance of the recommender system.

Project Risks

- **Data quality:** The data may be incomplete or inaccurate.
- **Management approach:** The results of an experiment can change the approach for modelling techniques used at the project. This may lead to inevitable changes in internal software architecture and major changes in the entire system
- **Machine learning algorithms:** The machine learning algorithms may not be able to learn the patterns in the data.
- **Performance:** The recommender system may not be able to accurately predict movies that users are likely to enjoy.

Project Techniques

- Popularity based Recommendation systems based on Weighted Rating
- Content Based Recommendation system based on NLP concept TF-IDF (term frequency - inverse document frequency)
- Collaborative Based Recommendation system based on KNN (K Nearest Neighbors)

Project Plan

- **Merging all the given datasets** (credits.csv, keywords.csv, links.csv, links_small.csv, movies_metadata.csv, ratings_small.csv)
 - credits and keywords, credits and movies_metadata on id
 - ratings_small and links on movieId
 - links and credits on tmdbId
- **Data cleaning**
- **Exploratory Data Analysis** (Data Visualisations)
- **Data Preprocessing**
- **Save the Dataframe** as csv file which will be used in Popularity, Content, Collaborative Based Recommendation systems
- **Model Building**
 - Weighted Rating for Popularity based Recommendation systems
 - TF-IDF (term frequency - inverse document frequency) for Content Based Recommendation system
 - KNN (K Nearest Neighbors) for Collaborative Based Recommendation system
- **Model Serialisation and DeSerialisation**
- **Deployment** using Streamlit where we have to select the type of recommendation system and give input accordingly to render output on screen

- **Report Generation** on Flow of project ,Difficulties faced , Ways to solve the difficulties, Future scope, Managing the app etc...

Project Team

- Project manager: @ **Koukuntla Komal Reddy**
- Data scientist: @ **Koppaka Yeswanth**
- Software engineer: @ **Chalumuri Harshitha**
- User experience designer: @ **Kalapala Venkata Surya Teja**
- Documentation writer:@ **Asif Ahmad Najar**
- Quality assurance engineer:@ **Chintada Jagan Mohan Rao**