

### Bachelor of Technology

### in

**COMPUTER SCIENCE AND ENGINEERING**

**22CS3503 – ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**

**MINI PROJECT REPORT**

On

### MOVIE SUCCESS PREDICTION

Submitted By

## SINCHANA K -ENG22CS0169

## SANGANA A-ENG22CS0150

## SNEHA-ENG22CS0173

## SNEHA M P-ENG22CS0176

### UNDER THE SUPERVISION

### Prof. Sasikala N

**Department of Computer Science and Engineering**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SCHOOL OF ENGINEERING**

**DAYANANDA SAGAR UNIVERSITY**

### (2024-2025)



**School of Engineering**

**Department of Computer Science & Engineering**

Harohalli, Ramanagara - 562112

Karnataka, India

**CERTIFICATE**

This is to certify that the Mini Project titled “**MOVIE SUCCESS PREDICTION**” carried out by **Sinchana k** **(ENG22CS0169), Sanjana (ENG22CS0150), Sneha (ENG22CS0173), Sneha M P (ENG22CS0176)** bonafide students of Bachelor of Technology in Computer Science and Engineering at the School of Engineering, Dayananda Sagar University, Bangalore in partial fulfillment for the award of degree in Bachelor of Technology in Computer Science and Engineering, during the year 2024-2025.

|  |  |  |
| --- | --- | --- |
| **Prof. Sasikala N** | **Dr. Girisha G S** | **Dr. Uday Kumar Reddy K R** |
| Asst Prof, Dept. of CSE  School of Engineering Dayananda Sagar University | Chairman, CSE  School of Engineering Dayananda Sagar University | Dean  School of Engineering  Dayananda Sagar University |
|  |  |  |

# **ACKNOWLEDGEMENT**

It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this AIML MINI PROJECT

First, we take this opportunity to express our sincere gratitude to the School of Engineering, Dayananda Sagar University for providing us with a great opportunity to pursue our bachelor’s degree in this institution.

We would like to thank **Dr. Uday Kumar Reddy K R, Dean**, **School of Engineering**, **Dayananda Sagar University** forhis constant encouragement and expert advice. It is immense pleasure to express our sincere thanks to **Dr. Girisha G S, Chairman**, **Department of Computer Science and Engineering**, **Dayananda Sagar University,** for providing the right academic guidance that made our task possible.

We would like to thank our teacher **Prof Sasikala N**, **Assistant Professor**, **Department of Computer Science and Engineering**, **Dayananda Sagar University**, for sparing her valuable time to extend help in every step of our R PROGRAMMING MINI PROJECT, which paved the way for smooth progress and the fruitful culmination of the project.

We are also grateful to our family and friends who provided us with every requirement throughout the course. We would like to thank one and all who directly or indirectly helped us in the R PROGRAMMING MINI PROJECT.

# **ABSTRACT**

Film Industry is not only a industry or a center of entertainment, rather it is now a center of global business. All over the world is now excited about a movie’s box office success, popularity etc. A huge data is available online about these movies’ success or popularity. The film industry is very important sector in the global market. Therefore, it is very necessary maximize the profit by predicting the movie success before its release. The number of movies released in the world is growing at an tremendous rate and success rate of movie is of very importance since million of dollars are invested in the making of each of these movies. In such a cases, prior knowledge about the success or failure of a movie and what factor affect a movie success will benefit the production houses since these predictions will give them a good idea of how to go about with the advertising , which itself is an expensive affair altogether. So, the prediction of the success of the movie is very necessary to the film industry. We proposed to develop a model for predicting the success of the movie being a Flop or Hit, long before a movie is actually released using machine learning techniques and algorithms.

## TABLE OF CONTENTS

**PAGE NO**

ABSTRACT

CHAPTER 1 INTRODUCTION………………………….………………………………………01

CHAPTER 2 PROJECT DESCRIPTION……………………….………………………………..02

CHAPTER 3 DESIGN…………………………….………………………………………………03

CHAPTER 4 METHODOLOGY……………….………………………………….…………….04

CHAPTER 5 SYSTEM IMPLEMENTATION.…….…..………………………………………………..05

CHAPTER 6 TESTING AND RESULT……...…………………………………………………………..07

CHAPTER 7 CONCLUSION…………..………………………………………………………. 08

CHAPTER 8 REFERENCES………………………...….…………………………………….... 09

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Fig. No.** | **Description** | **Page No.** |
|  |  |  |
|  |  |  |
|  |  |  |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Fig. No.** | **Description** | **Page No.** |
|  |  |  |
|  |  |  |
|  |  |  |

**CHAPTER 1**

**INTRODUCTION**

Now a days movies are not the only source of recreation, rather it is one of the major sources of global commerce and marketing. Movies create a new craze among people especially young people. Not only movie directors and box office officials are concerned with the success of the movie but general people also. People used to talk about these in social medias. Therefore, analysis of social media data about the movies is popular among the data analysts. And remains some other scopes like analyzing a director’s previous success histories or a actor’s previous popularity etc. Again, the analysis may be different on different countries. Naturally peoples from all the regions of the world do not react in the similar way. Movies are now available on internet. There are platforms like IMDb (Internet Movie Database) , Rotten Tomatoes , Meta critics etc. where people can share their reviews about movies. Movies continue to be a major source of entertainment in any country. However, this industry might faces many flops when the movie does not perform well at the Box Office. Our project will try to predict the movie success rate by doing predictive analysis on the many features of the movie. Day by day these platforms are becoming popular since people are getting honest reviews there. So, huge data is available online about reviews and ratings of movies. So, the prediction of the success of a movie is very essential to the film industry. We proposed to develop a model for predicting the success of movie being a flop or hit, long before a movie is actually released using machine learning techniques and algorithms.

**CHAPTER 2**

**PROJECT DESCRIPTION**

This project aims to predict the success of a movie using AI and machine learning techniques. Success is defined by factors like box office earnings, audience ratings, and critical reviews. The model will be trained on historical movie data, including features such as genre, cast, director, production budget, release date, and marketing efforts. By analyzing these factors, the model will identify patterns that contribute to a movie's success or failure. The goal is to help stakeholders make data-driven decisions in movie production, marketing, and investment. The project will use machine learning algorithms like regression, decision trees, and neural networks, with evaluation based on metrics such as accuracy, F1-score, and RMSE.

**CHAPTER 3**

**DESIGN**

The goal of this project is to predict a movie's success based on factors like budget, genre, cast, release date, and social media sentiment.

1. **Data Collection**: Gather data from IMDb, Box Office Mojo, Rotten Tomatoes, and social media sentiment.
2. **Data Preprocessing**: Clean and transform the data (handle missing values, encode categorical features, scale numerical ones).
3. **Modeling**: Use algorithms like Random Forest (for classification) and svm (for revenue prediction).
4. **Evaluation**: Assess model performance using metrics like accuracy or R-squared.
5. **Deployment**: Build an API with Flask to predict success based on user input.

6.**Technologies**: Python, Pandas.

**CHAPTER 4**

**METHODOLOGY**

**Data Preprocessing**: The machine learning technique plays major role in Data Preprocessing because it is the major part where the data would transform or get encoded in a way that all duplicate values would be deleted and also helps in shaping of the dataset.

**Importing Algorithms**: We have imported two algorithms in total they are, Support Vector Machine, and Random Forest. We have chosen only these particular algorithms because they are specifically used in classification process and as we have to classify the data based on which can be used for movie success prediction and which could not be used these classification algorithms helped in achieving the desired results.

**Performance Evaluation**: Our dataset was split in 80:20 ratio and was then considered for predicting the test results. The predicted results of data are tested for accuracy using several algorithms mentioned above and the algorithm with most accuracy was finalized for predicting whether the data could be used for movie success prediction

**User Model and Front End**: Our user model collects and would categorize data which are specific to user. This user model is made interactive with the help of front end which was done using html, CSS and JavaScript.

**Final Model**: Upon the results obtained by performance evaluation and from the input given by user using front end, Random Forest had 85.20% accuracy so we had finalized this model for predicting the output required.

**CHAPTER 5**

**SYSTEM IMPLEMENTATION**

**1.Data Collection**:

Sources: IMDb, Box Office Mojo, Rotten Tomatoes, social media (Twitter sentiment).

Features Collected:

Movie details: Genre, director, cast, budget, runtime, release date.

Sentiment data from reviews and social media.

Box office earnings (target variable), critical scores (IMDb, Rotten Tomatoes).

**2.Data Preprocessing:**

Data Cleaning: Handle missing values (impute or remove), remove duplicates.

Feature Engineering:

Convert categorical variables (e.g., genre, director) into numerical features (one-hot encoding).

Scale numerical features like budget, runtime, and box office earnings.

Sentiment Analysis: Use NLP techniques to extract sentiment from reviews and social media.

**3.Model Development**:

Split Data: Divide data into training and testing sets (80/20).

Model Selection:

For regression tasks (e.g., predicting box office revenue): Use Random Forests.

For classification tasks (e.g., predicting if a movie is a hit or flop): Use SVM

Train Model: Train the model on the training data, tune hyperparameters using Grid Search or Random Search.

Model Evaluation: Evaluate using appropriate metrics (R-squared for regression, Accuracy).

**Technologies Used**:

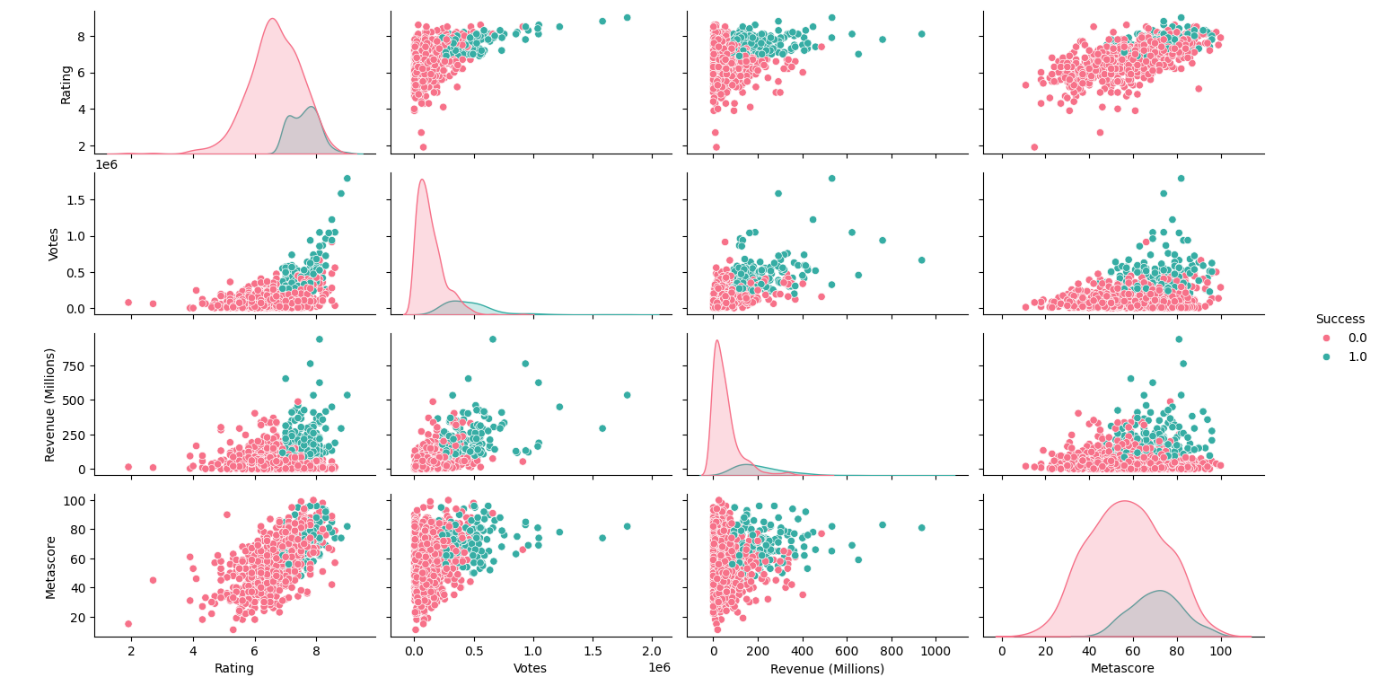
Programming Language: Python

Libraries: Pandas, NumPy

This system uses historical movie data, builds predictive models, and deploys them for real-time success predictions.

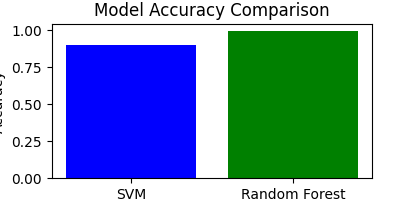
**CHAPTER 6**

**TESTING AND RESULT**



**FIG1.1**

**FIG 1.1**



**FIG1.2**

**CHAPTER 7**

**CONCLUSION**

Now a days movies are not the only source of recreation, rather it is one of the main sources of global commerce and marketing. Movies create a new craze among people especially young people. Not only movie directors and box office officials are concerned with the movie success but general people also. The prediction of success of movie with good accuracy is necessary in the film industry which helps different people working in the movie industry mainly for the investors. We proposed to develop a model for predicting the success of movie being a Flop or Hit, before the movie actually released using machine learning techniques and algorithms. accuracy of 85.20 %.

**CHAPTER 8**

**REFERNENCES**