**Movie Success Prediction using Data Science**

**A Project Report**

**Submitted By**

|  |  |  |
| --- | --- | --- |
| **Atreya Raorane** | **–** | **Class 74179** |
| **Mugdha Shinde** | **–** | **Class 74179** |
| **Yash Bhavsar** | **U01873220 –** | **Class 74179** |



**Guide Name:** **Prof. Stephan Barabasi**

**Introduction to Data Science**

**ACKNOWLEDGEMENT**

It is a matter of great honor and privilege for us to offer our grateful acknowledgment to our guide Prof. Stephan Barbasi of Pace University - Seidenberg School of Computer Science and Information Systems, New York for providing us an excellent chance to work under their guidance and supervision.

I would like to express my special thanks of gratitude to my esteemed guide, Prof. Stephan Barbasi who gave me the golden opportunity to do this wonderful project which also helped me in doing a lot of Research and I came to know about so many new things. I am thankful to them.

Finally, we would like to express our sincerest thanks to all the members of our family, who gives us strength and opportunity to aspire for this level of education.

# Abstract

In this data science project, we aim to predict the success of movies using a variety of factors such as budget, cast, genre, and release date. We will gather and clean a dataset of movie information and financial performance from a reliable source, such as Rotten Tomatoes.

Using machine learning algorithms, such as random forests or gradient boosting, we will fit a predictive model on the training data and evaluate its performance using various metrics

Finally, we will use the trained model to make predictions on the test set and interpret the results, providing insights that can inform decision-making in the entertainment industry.

We will then split the dataset into training and test sets and use a machine learning algorithm, such as random forests or gradient boosting, to fit a predictive model on the training data.

**INDEX**

|  |  |  |
| --- | --- | --- |
| *Chapter No.* | *Contents* | *Page No* |
|  | Acknowledgment | 2 |
|  | Abstract | 3 |
| **1.** | **Introduction** | **Page no** |
|  | Project Introduction | 5 |
|  | Project Objectives | 5 |
| 2. | System Requirement Study |  |
|  | Software Requirements | 6 |
|  | Jupiter Notebook | 6 |
|  | Tableau | 7 |
|  |
| 4. | Project Testing |  |
|  | Implementation and Output Screen of Project | 8 |
|  | Conclusion | 15 |
|  |  |  |

**INTRODUCTION**

In this project, we will be using data science techniques to predict the success of a movie based on various factors such as the budget, genre, cast, and more. We will be utilizing data from Rotten Tomatoes, and to gather the necessary information.

We will be using a variety of data science techniques such as data cleansing, feature engineering, and machine learning algorithms to analyze and understand the data. We will then use this understanding to develop a predictive model that can accurately predict the success of a movie based on the input data.

We hope that this project will provide valuable insights into the movie industry and help studios and filmmakers, as well as investors who want to invest in movies, make more informed decisions about their projects.

**OBJECTIVES**

The main objective of the "Movie Success Prediction Using Data Science" project is to develop a predictive model that can accurately predict the success of a movie based on various factors such as the budget, genre, cast, and more. This model will be developed using data science techniques such as data cleansing, and machine learning algorithms.

The ultimate goal of this project is to provide valuable insights into the movie industry and help studios and filmmakers make more informed decisions about their projects. By understanding which factors are most important in predicting a movie's success, studios and filmmakers can better allocate resources and make strategic decisions that are more likely to result in a successful movie. It will be useful for investors to invest in the movies.

Additionally, this project aims to demonstrate the power and utility of data science in predicting real-world outcomes. By using data science to analyze and understand complex data sets, we can gain a deeper understanding of the underlying patterns and trends that drive the success of a movie.

# Software Requirements

Software Requirement: -

1. Jupyter Notebook
2. Tableau
3. Jupyter Notebook:

Jupyter Notebook is an open-source web-based interactive computational environment for creating and sharing documents that contain live code, equations, visualizations, and narrative text. It allows you to create and share documents that contain live code, equations, visualizations, and explanatory text.

Jupyter Notebook is often used for data analysis, machine learning, and scientific computing. It is popular among data scientists and researchers because it provides an easy-to-use, interactive platform for analyzing and visualizing data, and for developing and testing machine learning models.

To use Jupyter Notebook, you will need to install it on your computer. You can install Jupyter Notebook using pip, the Python package manager. Once you have installed Jupyter Notebook, you can launch it by running the jupyter-notebook command in your terminal.

Jupyter Notebook consists of cells, which can contain text, code, or equations. You can execute code cells by selecting them and pressing Shift+Enter, or by clicking the Run button in the toolbar. You can also add new cells, delete cells, and move cells up or down in the notebook.

Jupyter Notebook supports a wide range of programming languages, including Python, R, Julia, and many others. It also includes support for interactive data visualization, so you can create plots and charts directly in your notebook.

Overall, Jupyter Notebook is a powerful tool for data analysis, scientific computing, and machine learning, and is widely used by researchers and data scientists around the world.

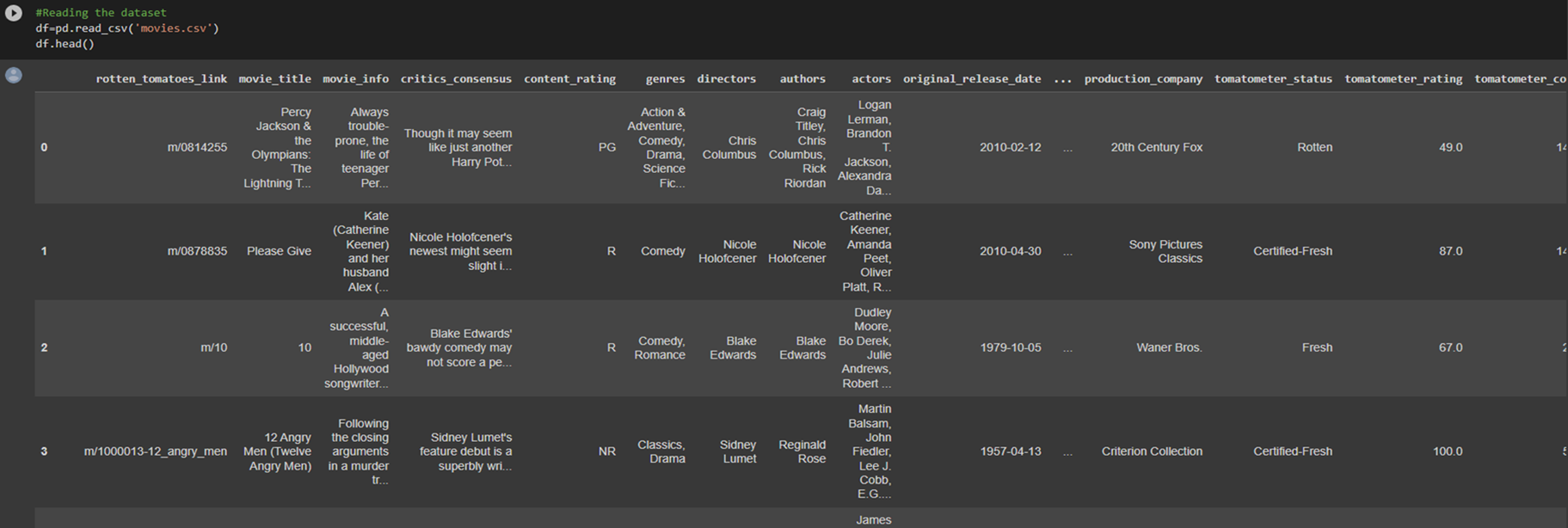
1. Tableau:

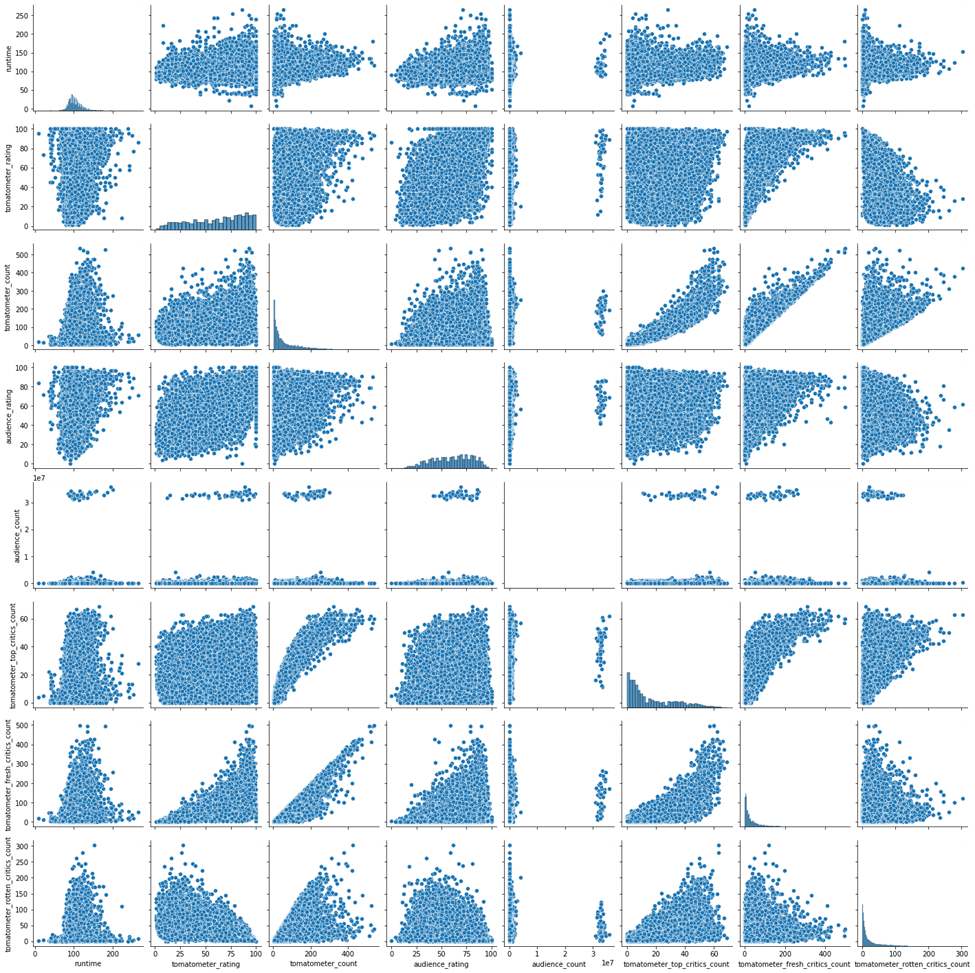
Tableau is a data visualization and business intelligence software that allows users to create interactive, shareable dashboards and reports. It is designed to help people see and understand data, and make data-driven decisions.

With Tableau, users can connect to a variety of data sources, including Excel, SQL databases, and cloud-based data sources, and then use a drag-and-drop interface to build charts, graphs, and other types of visualizations. These visualizations can be combined into dashboards and reports that can be shared with others in an organization.

Tableau offers a range of features and tools that make it easy to analyze and understand data, including the ability to drill down into details, create calculated fields, and forecast future trends. It also offers collaboration features that allow multiple users to work on the same dashboards and reports, and it integrates with other business intelligence and data management tools.

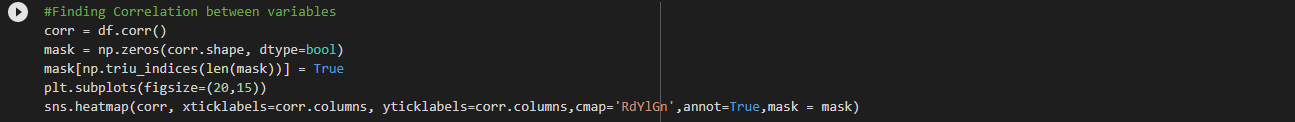
# Implementation & Output Screen of Project





Pair plot using seaborn library

Plotting correlation between variables using heatmap correlation graph



Chart

Description automatically generated

Chart, bar chart

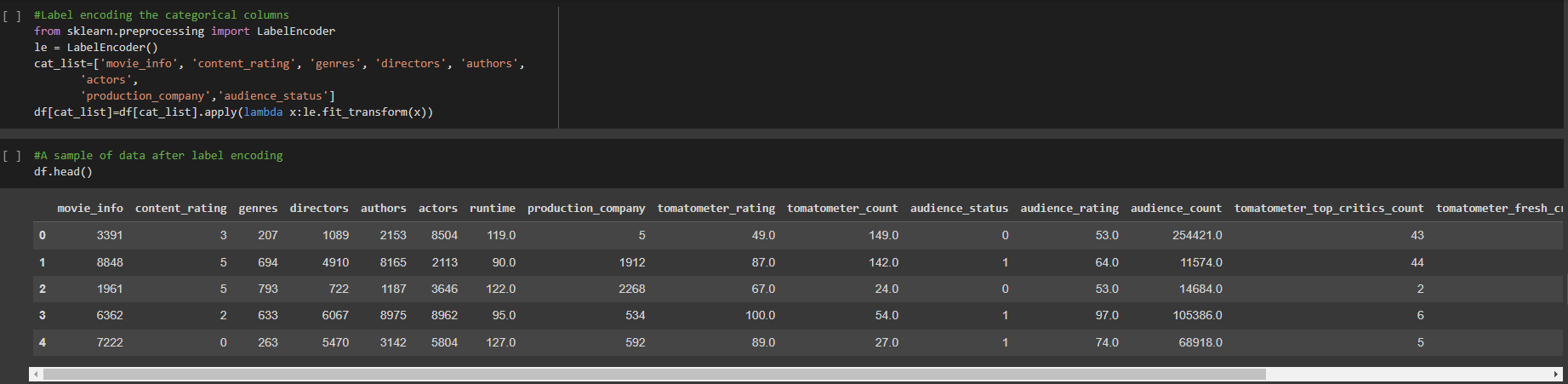
Description automatically generated

Chart, box and whisker chart

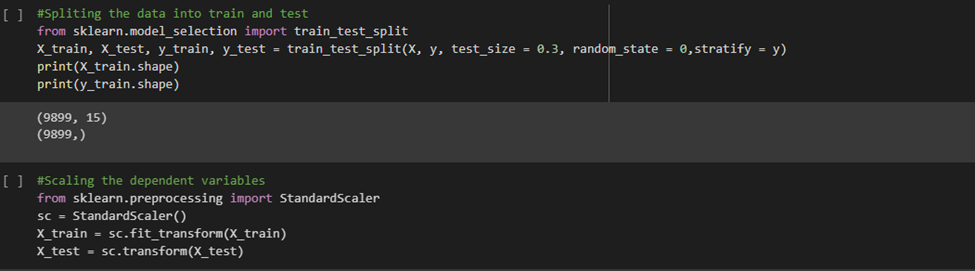
Description automatically generated

Bar graph and Box plot

Label Encoding of the data



Splitting The Data Into Training ,Testing Datasets and Scaling It.



Training and Testing the Data using Random Forest Classifier Algorithm

Graphical user interface, text

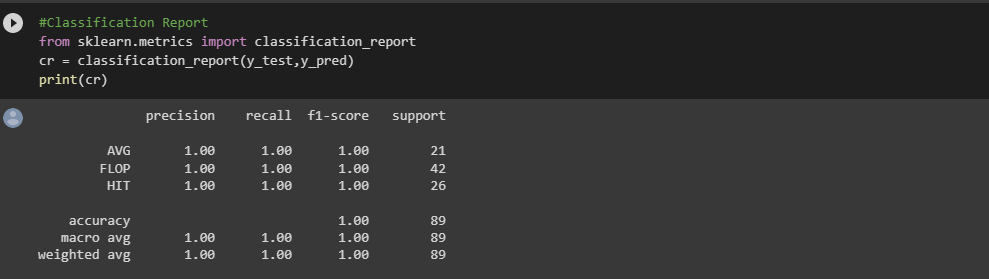
Description automatically generatedA picture containing text

Description automatically generated

Other Algorithms that we have tried:-

|  |  |
| --- | --- |
| Algorithms | Accuracy |
| KNN | 38 |
| Logistic Regression | 48 |
| Logistic Regression | 98 |

Trained Last 10 Years Data and Predicted Sci-Fie movies



**Data Visualization using Tableau**

Chart

Description automatically generated

Chart

Description automatically generated

# Conclusion

We were able to predict movie success rates by using different types of algorithms and exploratory data analysis with the help of different tools and libraries such as tableau and matplotlib and seaborn. By trying various algorithms, we came to know that random forest and Decision tree give us best accuracy while prediction .as well as we were able to predict only sci-fie movies by training last ten years data with accuracy of 100