SANTA CLARA UNIVERSITY

Department of Computer Science & Engineering

I HEREBY RECOMMEND THAT THE THESIS PREPARED
UNDER MY SUPERVISION BY

Caitlin Lopez, Kulsoom Sabit

ENTITLED

FILM VIABILITY EVALUATOR

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

BACHELOR OF SCIENCE

IN

COMPUTER SCIENCE & ENGINEERING

Yi Fang
Yi Fang (Jun 20, 2023 14:59 PDT) Jun 20, 2023

Thesis Advisor date

Silvia Figueira (Jun 21, 2023 21:41 PDT) Jun 21, 2023

Department Chair date

FILM VIABILITY EVALUATOR

By
Caitlin Lopez, Kulsoom Sabit

SENIOR DESIGN PROJECT REPORT

Submitted to
the Department of Computer Science & Engineering

of

SANTA CLARA UNIVERSITY

in Partial Fulfillment of the Requirements
for the degree of
Bachelor of Science in Computer Science & Engineering

Santa Clara, California

2023

Film Viability Evaluator

Caitlin Lopez, Kulsoom Sabit

Department of Computer Science & Engineering
Santa Clara University
2023

ABSTRACT

Filmmaking is a financially high-risk field of work. There are multiple variables that contribute to a movie's success or failure, however the most impactful ones have yet to be identified. In this paper, we used publicly available datasets in combination with several forms of analysis models to determine common factors associated with historically successful movies. Our results narrowed the top categories to include the runtime, the production budget, the release year, the genres, and the primary language. The final round of analysis yielded an 84% accuracy, after the usage of XGBoosting and gradient boosting. Further improvements to this project could involve testing different modeling techniques to improve accuracy, as well as using an expanded dataset that includes more films from outside of the United States.

Keywords: Filmmaking, Data Modeling, Financial Success

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST of TABLES

Page

Chapter 1 - Introduction

One of the most ubiquitous aspects of many peoples lives are movies. They are a universally accepted and loved language where anyone can find common ground. Everybody has some movie that has influenced them in their life, from the way they make decisions to the way that they lead the lives they live. However, with the entry of COVID-19, film industries all across the world witnessed a significant decline in sales and business. Once the pandemic got to a point where it was feasible to return to theatres, unfortunately people did not return in the numbers to revive the industry. In a Variety article from the summer of 2022, they detailed:

> *A record-high 88% of moviegoers are "very or somewhat comfortable" going to the movies, according to the latest study from the National Research Group. Around a year ago, that percentage was closer to 59%. That means people are feeling better and better about returning to a darkened room with strangers, a reality that prevented many ticket buyers from going to the movies in the earlier days of the pandemic. (Rubin)*

Film industries everywhere continue to struggle up until now as they continue to put out movies in hopes to return to the success of previous years, but have difficulty figuring out what audiences want.

Project Goal

The goal of this project was to create a product that people in the film industry can use as a frame of reference in order to create movies that have the capabilities of financial success. We wanted to find an intersection between film and education when pursuing this project. Given that the scope of the film industry is so large, it was important to optimize the scope of this project to a point where the imperative roles that go into the production of a film can be used towards this project. We have created a product called the Film Viability Evaluator, which takes into account data from over 5,000 movies spanning over 100 years. Using machine learning techniques, we have created a platform so that people in different roles can reference different categories that go into creating a movie, providing them with data that guides them based on the factors that they choose. The idea for the project is to provide a sense of help to producers, writers, directors, and students who choose to take advantage its benefits.

Use Cases

There are several potential users that would benefit from our product. Ranging from the film industry itself to educational spaces, many applications would be possible. Although there are many different use cases, the three primary ones we chose to focus on were producers, writers, and education.

Producers

One use case would be producers, both individual supporters and production companies. Using the analyzed data, producers can more easily decide which projects to provide funding to because they will know which attributes yield the most successful results. They can also use the evaluator to propose new ideas to writers and directors, providing reliable data to back up their proposals. Currently, many production companies having only been focusing on funding remakes or continuations of movie franchises because they are afraid to take the financial risks on new stories. With our product, filmmaking can hopefully go in the other direction in order to amplify more diverse voices in popular media.

Writers

Secondly, writers themselves would benefit from using the evaluator, specifically if they are planning to pitch their screenplay. They can compare their existing work to the historical data and make revisions to things such as the length of the script or languages spoken by the characters, increasing the chances of their story getting produced. At the time of writing this paper, the Writers Guild of America, a labor union representing writers in film, television, radio, and online media, has been on strike, advocating for proper compensation for their work (WGA). This in particular emphasizes the importance of writers needing the ability to have more control over their creative work and its outcomes.

Education

Another use case would be within education, both for educators and students. In regards to educators, the evaluator would help them to teach better filmmaking practices, as they would have better guidelines to recommend to students. For students, the tool would hopefully give them a better foundation in filmmaking, helping them determine which "rules" to follow and which ones they can challenge.

Chapter 2 - Overall System Integration

System Overview

Our product is a web-based portal that generates a set of analysis models based on historic film data and recommends feasible criteria when producing a new film. The home page displays a selection of successful movies that are a part of the dataset. This dataset was publicly available, using information from IMDb.

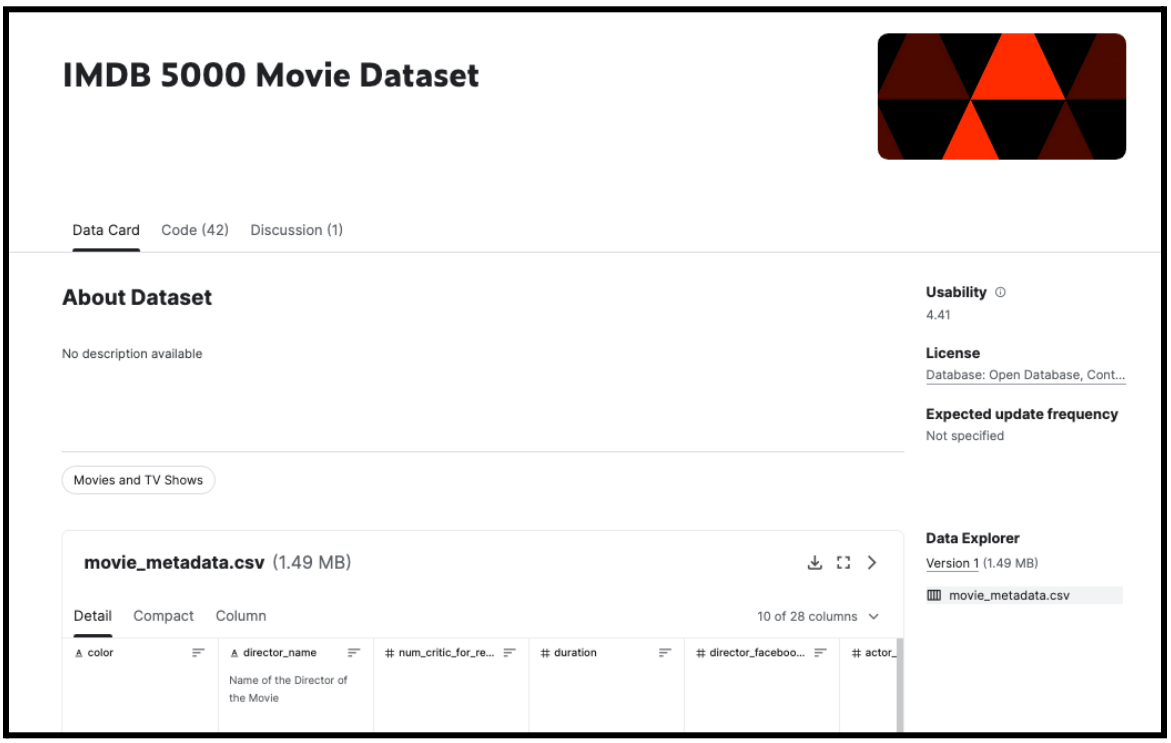


Figure 2-1 Data source.

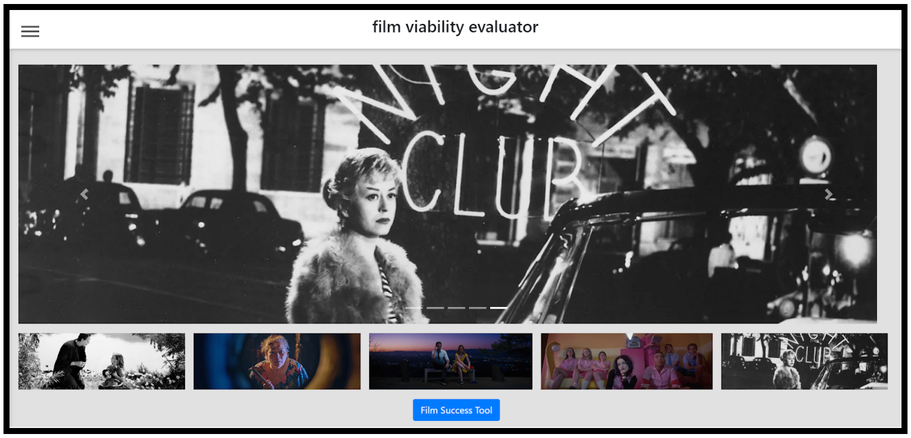


Figure 2-2 Home page.

Clicking on any of these thumbnails on the home page leads to an individual profile of each film's data that corresponds to the evaluator. We decided to make this design choice for new users of the website to have a better experience while using the website: rather than just having the evaluator as the first thing, users can ease into the content and explore before diving in.

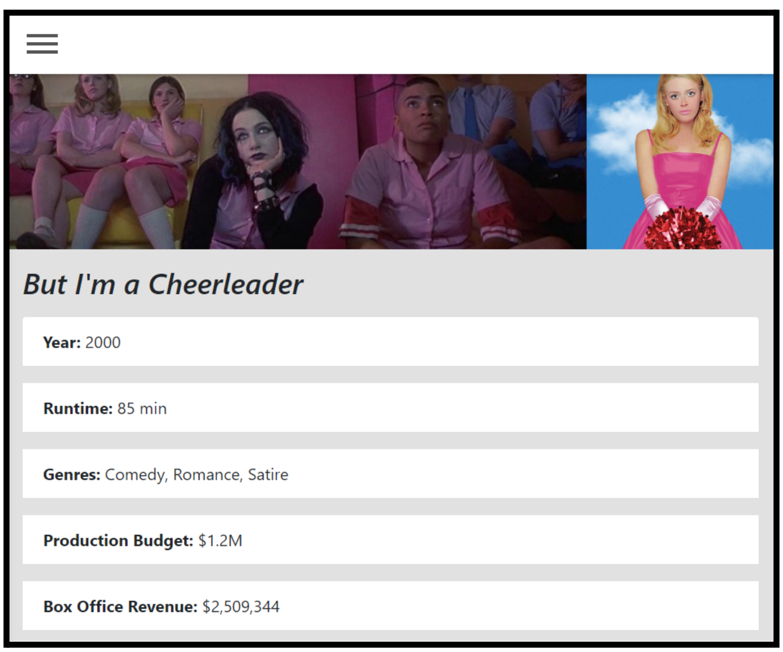


Figure 2-3 "But I'm a Cheerleader" data.

Moving on to the evaluator tool itself, users are first led to a page asking them which user category they fall under. We wanted to do this in order to better personalize the user experience, taking out information that might not be as applicable to their needs and would end up being overwhelming if included. This is important to think about as engineers because having unnecessary content both makes the product less effective for users and also requires more resources to be spent (in our case memory and time).

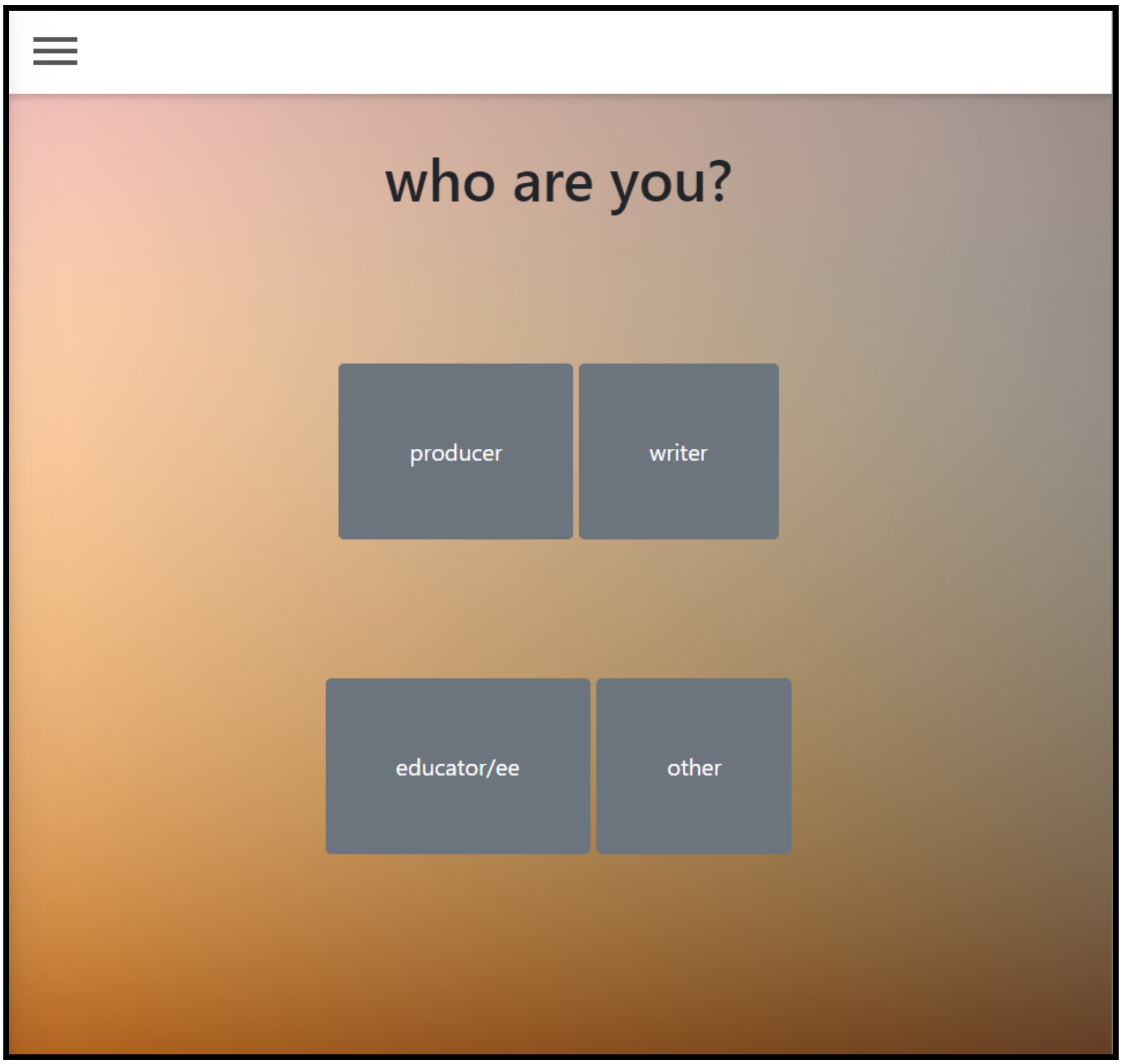


Figure 2-4 User category options.

After selecting an option, the user is directed to categorized data tailored to their career needs, each option containing specific data that is more useful to them. Below shows the tool specific for writers. Users can view all data categories in order of importance or filter specific categories they would like to see.

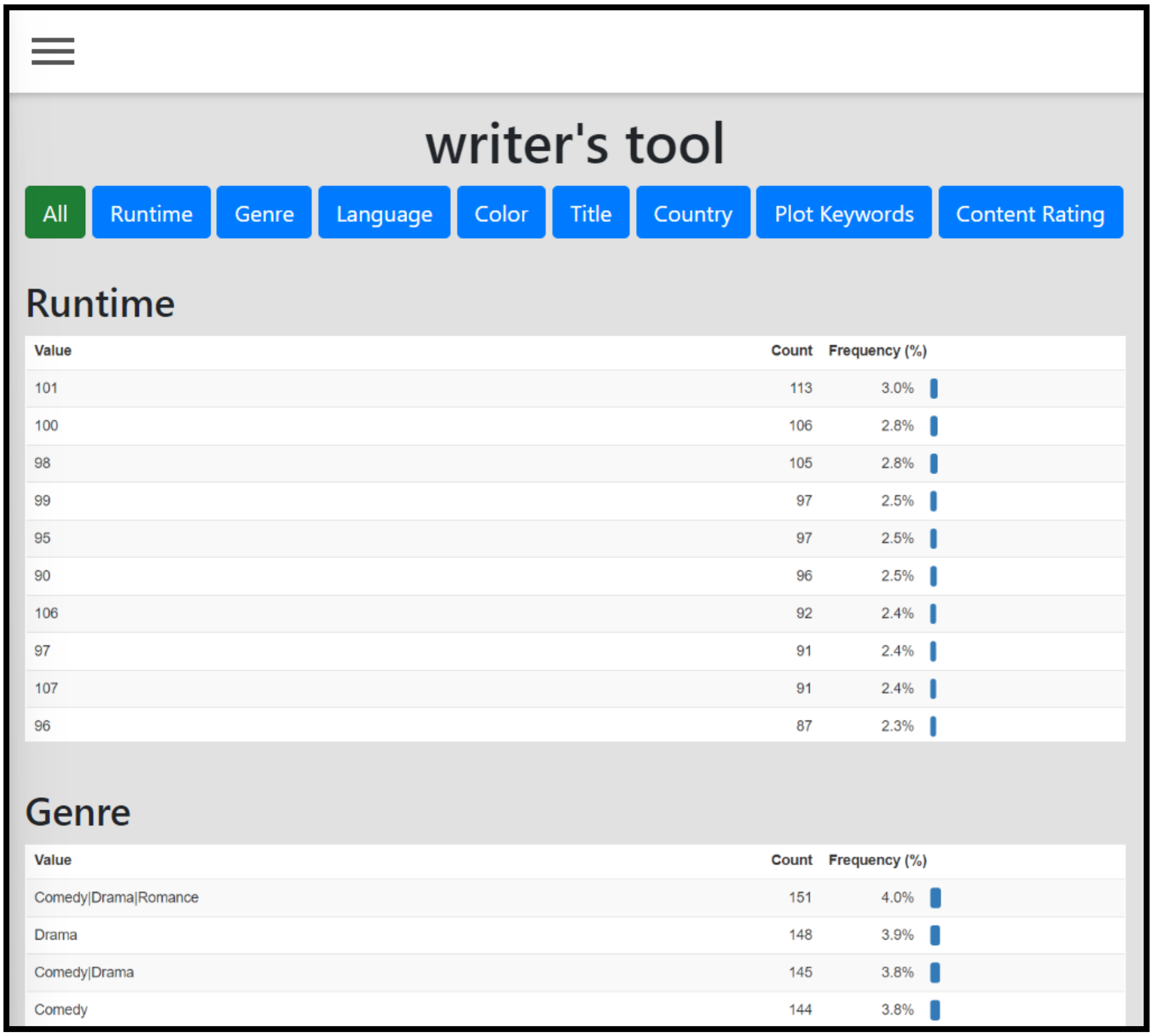


Figure 2-5 Writer's tool.

Team Structure

For this project, we worked as a two-person team. Although this allowed scheduling meetings to be relatively easy, this meant there was a lot more work to be divided among ourselves. Kulsoom had some experience with machine learning when we first began this project, and Caitlin had some experience with artificial intelligence and a couple years of background on web development.

Design Process

Our design process began with identifying the exact requirements we wanted our product to adhere to. First, we looked to three different professional standards within computer science and engineering, one being the AIS Standards outlined by IEEE. Specifically, the "Standard Model Process for Addressing Ethical Concerns during System Design" and "Algorithmic Bias Considerations". Another set is the HTML Standard, making sure we follow good practices when building the website portion. The third set are the guidelines defined to comply with ADA, so the tool is useful for everybody.

Next, we chose personal requirements, both functional and nonfunctional. The functional requirements for this project are that the program is able to process our datasets and produce useful models, and the website containing this tool is reliable and has little to no issues when handling user actions. The nonfunctional requirements for this project include ensuring the usability of the interface, optimizing the performance of website in conjunction with the ML models, and making sure that the product is viable.

Another important part of the design was taking into account potential obstacles we could encounter during the course of the project. To prepare for this, we completed a risk analysis table.

Table 2-1 Risk analysis

Scenario	Risk	Severity	Likelihood	Mitigation Strategy
Group member illness	72	9	8	assigning backup roles
Task dependency	90	10	9	finalize main components before working on additional features
Additional research needed	25	5	5	read into technologies as much as possible before using
Errors within code	80	8	10	double check code for common mistakes (typos, missing punctuation)
Scheduling conflicts	36	6	6	set meetings earlier than needed
More time needed than expected	49	7	7	building in buffer time

To calculate the risk level of each scenario, we used the equation risk = severity * likelihood, all of the input values using a scale from 1 to 10. From there, we were able to rank the risk levels from most critical to least:

1. Task dependency (90)
2. Errors within code (80)
3. Group member illness (72)
4. More time needed than expected (49)
5. Scheduling conflicts (36)
6. Additional research needed (25)

Following this, we created a project timeline to adhere to throughout the school year, which can be found in Appendix A. We made sure to provide ourselves with some buffer to give us enough time to complete every stage of the project. Unfortunately, we ran into all of the scenarios in our risk analysis chart over the year, so the pacing of our work flow was not as smooth as we had hoped for.

The next step was to plan out our code structure using UML diagrams. We decided to make one that detailed the backend and then a separate one that connects the backend to the front end.

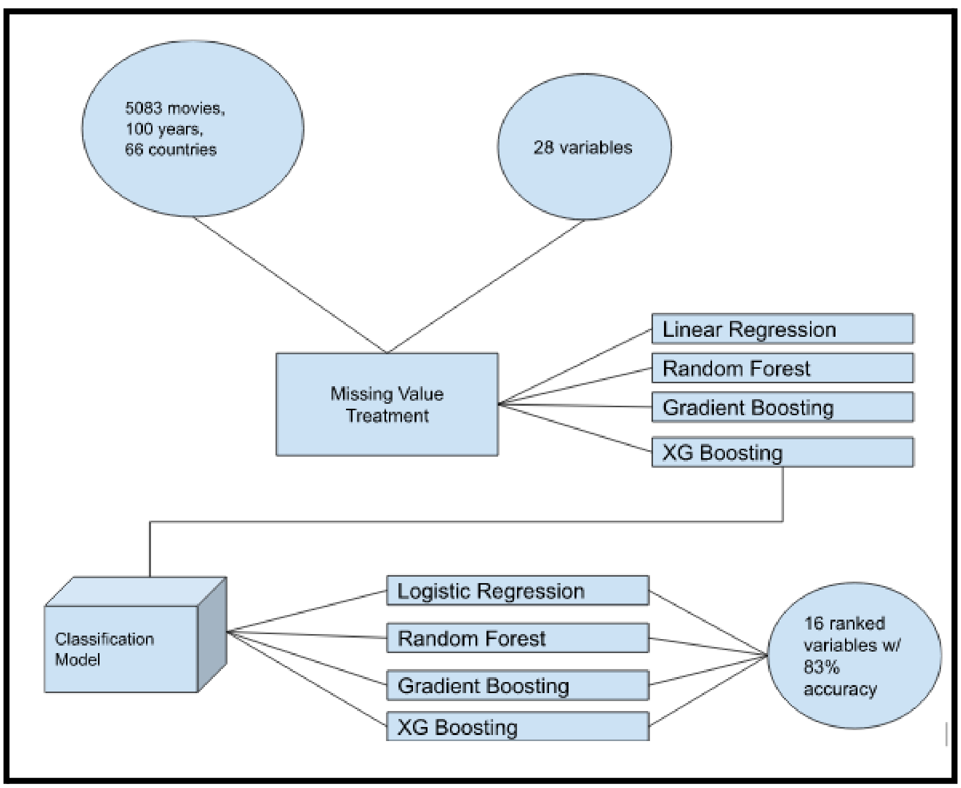


Figure 2-5 Machine learning architecture diagram.

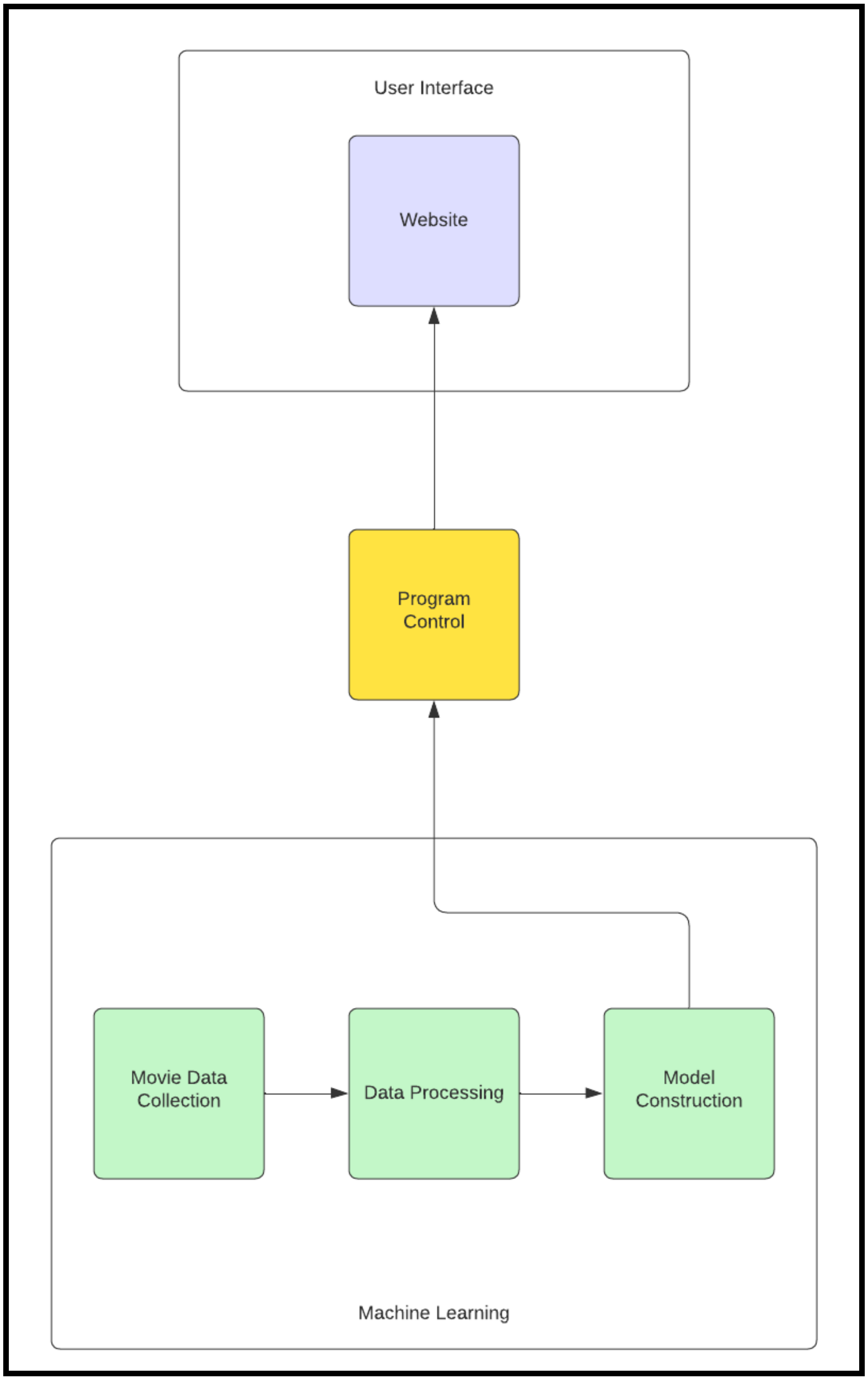


Figure 2-7 System architecture diagram.

Technologies Used

We decided the best language to use for this project would be Python since a majority or machine learning programs and implementations are done in the language. Given that regression techniques are also fairly complicated, Python would be the best bet in implementing these algorithms to simplify and organize things. One machine library we implemented is NumPy in order to deal with regression techniques, matrices. We also used pandas to convert to NumPy values. Regression techniques were necessary for parsing through and analyzing the datasets in our project. The databases we drew our datasets from include data from a website called Kaggle, which stores script and film datasets based on genre, runtime, box office collection, etc. This was our primary dataset source as it has an extensive database of data we needed for this project.

Framework wise, alot of our code will be comprised of clustering techniques, linear and ridge regression, LSTM, RSME, and alot of parsing. We began with clustering and scrapping of the data in order to isolate our datasets into just the categories we used like genre, runtime, and box office and removed categories like actors used, movie codes, etc. Following this, we implemented Linear Regression in order to create our prediction models per category and followed that with Ridge Regression because of expected cases of underfitting or overfitting. Finally, we used LSTM in order to classify our sequential data as its been analyzed; providing us with long term dependencies given the data we analyzed.

Chapter 3 - Modeling Techniques

There were numerous machine learning modeling techniques that were implemented to program the machine learning aspect of this project. The main language used for this part was Python and the project was split into two phases of creation. Phase one was primarily made up of data cleaning in order to scale down the data set to a point that it was feasible for statistic analysis. Given that the data set contains over 5,000 movies along with 28 different categories, it was necessary to prioritize certain categories over others and this included scaling it down from those that had empty data or missing values.

The data set that was used spans over 100 years and 66 countries holding over 5,000 movies. This is a very large data set and is not easy to work with when it comes to creating the project, so phase one on the majority was used to clean the data set and scale it. Following this, it was decided that first a basic regression model would be implemented and then it would advance to an advanced classification model which would implement these statistical analyses that we needed. A lot of the values in the data set held data that was not completely necessary towards the project or had missing values, for example the genre category would held 500 values whereas language would only hold 100. Because of comparison of categories like this, it was necessary to implement missing value treatment to the data set, which on a basic level takes the data and cleans it to prioritize those categories that have closer number values to each other rather than those that have few values or are completely empty. When cleaning the data and implementing this small regression model, the analysis was giving us a lot of errors and warning messages letting us know that our data was very messy and there were values that were skewing the entire model to a place that did not fit the regression line that was produced.

Dataset info

Number of variables	28
Number of observations	3817
Missing cells	0 (0.0%)
Duplicate rows	0 (0.0%)
Total size in memory	835.1 KiB
Average record size in memory	224.0 B

Variables types

Numeric	15
Categorical	11
Boolean	0
Date	0
URL	1
Text (Unique)	0
Rejected	1
Unsupported	0

Warnings

`actor_1_facebook_likes` is highly skewed (γ1 = 20.28118749)	Skewed
`actor_1_name` has a high cardinality: 1485 distinct values	Warning
`actor_2_name` has a high cardinality: 2259 distinct values	Warning
`actor_3_name` has a high cardinality: 2661 distinct values	Warning
`budget` is highly skewed (γ1 = 44.40420967)	Skewed
`cast_total_facebook_likes` is highly correlated with `actor_1_facebook_likes` (ρ = 0.9464018081)	Rejected
`director_facebook_likes` has 661 (17.3%) zeros	Zeros
`director_name` has a high cardinality: 1723 distinct values	Warning
`facenumber_in_poster` has 1611 (42.2%) zeros	Zeros
`genres` has a high cardinality: 758 distinct values	Warning

48]:

Dealing with Null Data amount we have lost 25% of the given data. Let's deal with converting the Data in to numericals to feed our model.

Figure 3-1 Missing value treatment result.

Above shows the result of the data set right after applying an initial missing value treatment and cleaning which provides the us with a look at how there are warnings about categories that have high cardinalities, a lot of distinct values, or simply many zeros meaning missing data. The goal following this output was to clean the data even more and scale down categories, so instead of having 28 categories it would come down to however many were necessary towards the project with enough data to work with.

Following the plan to continue the regression model into a more advanced approach, we applied a linear regression to our model. At this point we had scaled down the categories very minimally, removing any values that have empty data or ones that have a lower number of data compared to others. The goal now was to see how accurate our data is using the application of a simple linear regression, which resulted in an r-squared value of 0.37.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             imdb_score   R-squared:                       0.377
Model:                            OLS   Adj. R-squared:                  0.374
Method:                 Least Squares   F-statistic:                     122.4
Date:                Wed, 04 Sep 2019   Prob (F-statistic):          7.83e-298
Time:                        02:23:39   Log-Likelihood:                -3793.9
No. Observations:                3053   AIC:                             7620.
Df Residuals:                    3037   BIC:                             7716.
Df Model:                          15
Covariance Type:            nonrobust
```

Figure 3-2 Linear regression results.

This value meant that our data was not very accurate, as the perfect model has an r-squared value of one. We then moved into the application of ensemble modeling. Three different types of modeling techniques were applied to the current data set in order to make it more accurate and precise, along with scaling down our categories in order to prioritize the ones that would contribute to the project the most. There was the application of gradient boosting which is known as one of the strongest boosting techniques, as it provides a lot of precision. Gradient boosting gave us an error rate of 0.42.

After this, we applied a modeling technique that is a branch of Gradient Boosting called Random Forest boosting. The problem with gradient boosting is that the larger the data set, the slower gradient boosting is applied, so as a result Random Forest is much faster and is also used to follow the goal of having a lower error rate. This was achieved within this context because the Random Forest boosting error rate gave a result of 0.41. Lastly, the Random Forest had greater precision and lower error rate when applying Random Forest with XG Boosting, which is a second branch of Random Forest and follows the same intent to provide a lower error rate along with a larger data set. As a result, the final model's error rate was 0.40 along with the categories being scaled down to the sixteen primary categories that were prioritized over the original 28.

```
color
num_critic_for_reviews
duration
actor_3_facebook_likes
actor_1_facebook_likes
gross
num_voted_users
facenumber_in_poster
num_user_for_reviews
language
budget
title_year
actor_2_facebook_likes
movie_facebook_likes
main_genre
```

Figure 3-3 Primary data categories.

Now that a basic model is established in scaled down categories, it was important to apply a more advanced classification model in order to test the accuracy of the data and be able to continue on to building the front end of the project. An advanced classification model would require that we use the same scale data that resulted from phase two with the same modeling techniques but a new target variable. This would happen in the form of a logistic regression model, which is also called a classic binary model. The point of this was to take our data and results in a line which fit because the linear regression model resulted in data that did not fit within the line that was output. In phases one and two there was a linear regression applied with gradient boosting, Random Forest boosting, and random forest with XG Boosting, whereas now logistic regression would be applied followed by the same three modeling techniques with new target variables. With this we were able to extract the accuracy rates of 84% for Random Forest and 83% for Gradient Boosting. At this point the project was scaled down from 20 categories to 16 along with a 84% accuracy rate based on prioritized categories shown below.

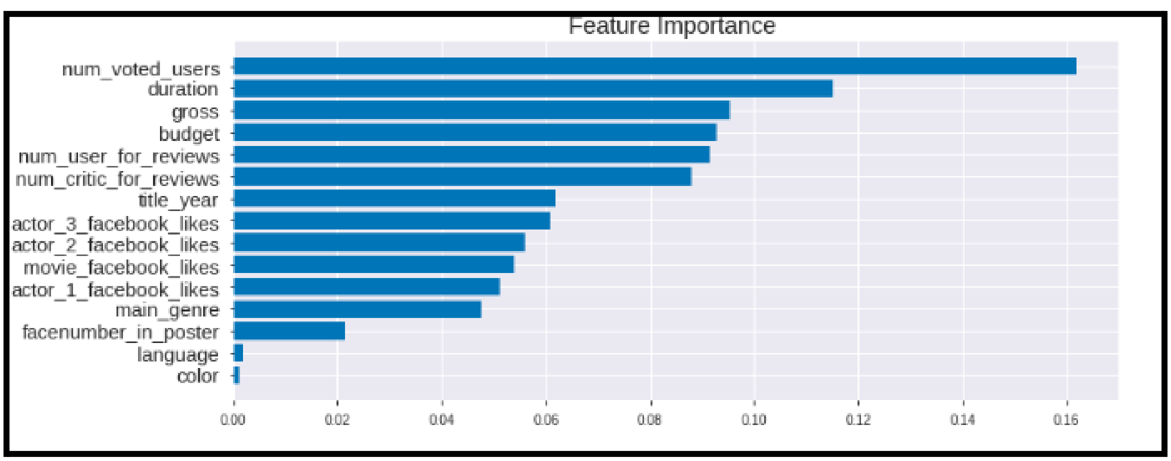


Figure 3-4 Prioritized film data.

Chapter 4 - Usability

User Experience

A positive user experience was one of the important objectives we wanted to achieve within our project. The interface plays a big role in this, so we outlined three key components we aimed for our UI to have: responsiveness, simplicity, and consistency. These aspects pointed us towards implementing the Bootstrap framework in our design, as it made development for both mobile and desktop devices easier and provided a great foundation for the layout of our website.

Environmental Impact

Our product has the potential to make a positive impact in terms of sustainability. Filmmaking is often a process with a heavy carbon footprint:

> *Blockbuster films with budgets of over $70 million produce an average of 2,840 ton of CO2 per production (it takes 3,700 acres of forest to absorb the equivalent in a year). Often filmed across several countries, 51% of these vast CO2 emissions are transport related.*
>
> *On a production set, several copies of the all-important script are printed on paper and throwaway plastic is often used to keep the cast and crew fed and hydrated. Props and costumes are often discarded after use, and huge convoys of vehicles and power generators leave a substantial carbon footprint. (Whittington)*

With the use of our evaluator, a lot of this waste can be mitigated. With more concentrated plans in terms of budget, the amount of things consumed during the production of a movie, whether that be food, electricity, or props, are reduced. Something like costumes may require more intentional decisions to stay within budget, such as buying pieces from thrift stores as opposed to fast fashion brands.

Chapter 5 - Ethical Evaluation

As for any software and technical product it is important to take into account the ethical considerations for the product being developed. Oftentimes there is a disconnect between people being able to see that there are ethical concerns when it comes to creating technology and an increasingly technological developing world. It is important to create ethical evaluations of our product in the context of these issues. The first and most one is software engineering. In the act of software engineering, especially when implementing machine learning techniques that could be used for artificial intelligence, it is imperative to take into account that there can be a huge bias when it comes to machine learning implementation. Software engineering is a technique that's relied on to provide real statistical data without the bias of things in society like race or class, but these considerations had to be taken into account within the context of our project because we are using historical data and historical data always comes with a sense of bias. It could be that specific films didn't do well because they contained actors of a specific ethnicity, or it could also be that one movie made a specific amount of money over another movie simply because of the language or region that it was created in. These categories and the results are all being implemented into the software engineering aspect of our project which needs to be accounted for.

Another ethical consideration to take into account in the context of software engineering is security and privacy. Our product uses a historical data set of over 5,000 movies and although the data set comes from an outside source, it isn't confirmed that the writers and directors of these movies have given permission for their creative work to be used within a data set like this and for projects that use the data. These are very serious ethical concern to take into account especially when catering a project to creative work.

Chapter 6 - Economic Considerations

During our work on this project, we were able to utilize free resources to accomplish our goals. This included the public datasets we used as well as using GitHub to host our webpage. Even though we were able to not spend any money to reach our findings, if this project was scaled up to be readily available for public use we would need sources of funding. This would cover costs such as web hosting, larger and more complete data sets, and possibly user-input data storage. To pay for the cost of product development, we could go the common route of having ads on our site, or we could also apply for grants to further research this topic. The tool could also be offered at different pricepoints (free, monthly subscription), and each level would offer more services than the previous level(s).

Chapter 7 - Conclusion

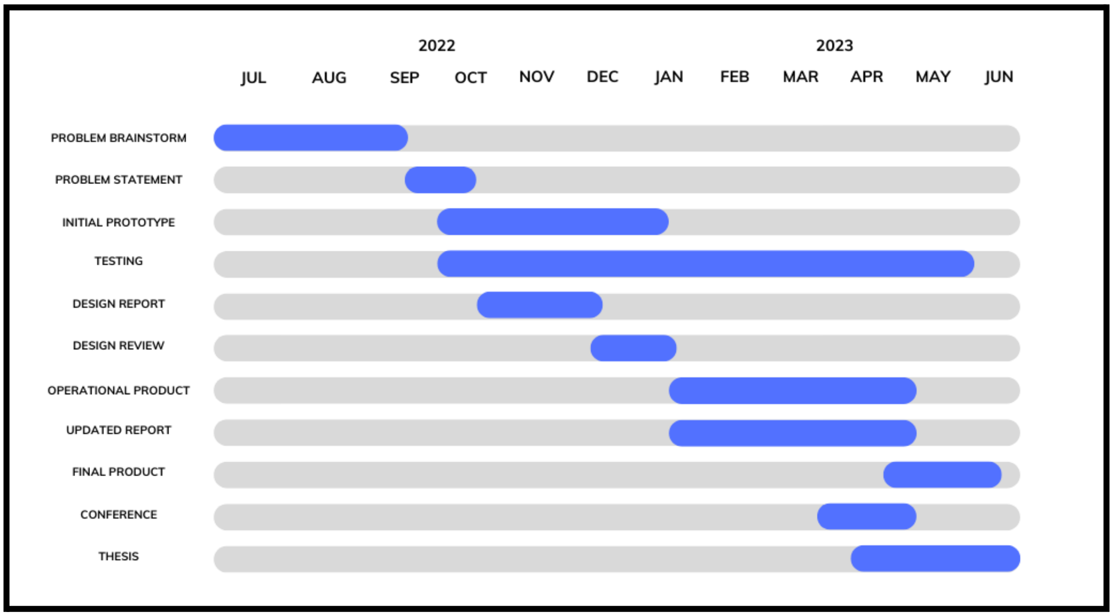
Summary

In a roundabout idea this project was intended to become a product that can be used by a people in a field (filmmaking) which we feel very passionate about. Our hard work and passion pushed towards a product that people in the film industry can use to create better movies for the general public. We implemented numerous machine learning techniques and also pursued a very high accuracy rate in order to attest to the idea that we created a product that is both accurate and useful. This was finally implemented with an interface in the form of a website that users can utilize when needed.

Future Uses and Further Development

There is a lot of room for growth within this product, especially with the number of categories being used in our original data set only being 5000 movies. Another aspect of growth and future use is that this product can be catered to specific countries, so if someone from India., for example, wanted to create a movie, then they can be provided statistical data on what works in India rather than in the United States or in Italy. An additional future development is extending the product to roles beyond producer, writer, and education, such as directors or cinematographers. Lastly, it can be more accessible by being turned into a mobile application which people can access on their phones much easier rather than having to access the website on a computer. The product at its current state is on a smaller scale and we would want to extend its functionality and make it better for the general public if we are able to further develop the project.

Appendix A: Timeline



Appendix B: Source Code

```
[37]  dataset.plot_keywords.unique()

...   array(['avatar|future|marine|native|paraplegic',
             'goddess|marriage ceremony|marriage proposal|pirate|singapore',
             'bomb|espionage|sequel|spy|terrorist', ...,
             'assassin|death|guitar|gun|mariachi',
             'written and directed by cast member',
             'actress name in title|crush|date|four word title|video camera'],
            dtype=object)
```

```
[38]  dataset.language.unique()

...   array(['English', 'Mandarin', 'Aboriginal', 'Spanish', 'French',
             'Filipino', 'Maya', 'Kazakh', 'Telugu', 'Cantonese', 'Japanese',
             'Aramaic', 'Italian', 'Dutch', 'Dari', 'German', 'Mongolian',
             'Thai', 'Bosnian', 'Korean', 'Hungarian', 'Hindi', nan,
             'Icelandic', 'Danish', 'Portuguese', 'Norwegian', 'Czech',
             'Russian', 'None', 'Zulu', 'Hebrew', 'Dzongkha', 'Arabic',
             'Vietnamese', 'Indonesian', 'Romanian', 'Persian', 'Swedish'],
            dtype=object)
```

```
[39]  dataset.language.value_counts()

...   English       3665
      French          37
      Spanish         26
      Mandarin        14
      German          13
      Japanese        12
      Hindi           10
      Cantonese        8
      Italian          7
      Korean           5
      Portuguese       5
      Norwegian        4
      Danish           3
      Thai             3
```

Python sample, before data cleaning.

```
from sklearn import ensemble
n_trees=200
gradientboost = ensemble.GradientBoostingRegressor(loss='ls',learning_rate=0.03,n_estimators=n_trees,max_depth=4)
gradientboost.fit(X_train_rfe,y_train)
```

[136]

```
GradientBoostingRegressor(alpha=0.9, criterion='friedman_mse', init=None,
                          learning_rate=0.03, loss='ls', max_depth=4,
                          max_features=None, max_leaf_nodes=None,
                          min_impurity_decrease=0.0, min_impurity_split=None,
                          min_samples_leaf=1, min_samples_split=2,
                          min_weight_fraction_leaf=0.0, n_estimators=200,
                          n_iter_no_change=None, presort='auto',
                          random_state=None, subsample=1.0, tol=0.0001,
                          validation_fraction=0.1, verbose=0, warm_start=False)
```

```
y_pred_gb=gradientboost.predict(X_test_rfe)
error=gradientboost.loss_(y_test,y_pred_gb) ##Loss function== Mean square error
print("MSE:%.3f" % error)
```

[137]

```
MSE:0.444
```

```
mean_squared_error(y_pred_gb, y_test)
```

[138]

```
0.4442879667702143
```

```
y_pred_gb.min(), y_pred_gb.max()
```

[139]

```
(4.249968514984295, 8.620023088931495)
```

```
from sklearn.model_selection import GridSearchCV
# Create the parameter grid based on the results of random search
param_grid = {
    'loss' : ['ls'],
```

Python sample, data cleaning.

```html
        <!--carousel-->
        <br>
        <div id="demo" class="carousel slide car" data-ride="carousel">
            <ul class="carousel-indicators">
                <li data-target="#demo" data-slide-to="0" class="active"></li>
                <li data-target="#demo" data-slide-to="1"></li>
                <li data-target="#demo" data-slide-to="2"></li>
                <li data-target="#demo" data-slide-to="3"></li>
                <li data-target="#demo" data-slide-to="4"></li>
            </ul>

            <div class="carousel-inner">
                <div class="carousel-item active">
                    <a href="films/frankenstein.html"><img src="img/frankenstein.jpg" alt="Image 1"></a>
                </div>
                <div class="carousel-item">
                    <a href="films/leonor.html"><img src="img/leonorwillneverdie.jpg" alt="Image 2"></a>
                </div>
                <div class="carousel-item">
                    <a href="films/lalaland.html"><img src="img/lalaland.jpg" alt="Image 3"></a>
                </div>
                <div class="carousel-item">
                    <a href="films/biac.html"><img src="img/butimacheerleader.jpg" alt="Image 4"></a>
                </div>
                <div class="carousel-item">
                    <a href="films/cabiria.html"><img src="img/nightsofcabiria.jpg" alt="Image 5"></a>
                </div>
            </div>

            <a class="carousel-control-prev arrow" href="#demo" data-slide="prev">
                <span class="carousel-control-prev-icon"></span>
            </a>

            <a class="carousel-control-next arrow" href="#demo" data-slide="next">
                <span class="carousel-control-next-icon"></span>
            </a>
        </div>

        <!--under carousel-->
        <div class="carousel-preview mt-3">
            <img src="img/frankenstein.jpg" alt="Image 1" data-target="#demo" data-slide-to="0" class="active">
            <img src="img/leonorwillneverdie.jpg" alt="Image 2" data-target="#demo" data-slide-to="1">
            <img src="img/lalaland.jpg" alt="Image 3" data-target="#demo" data-slide-to="2">
            <img src="img/butimacheerleader.jpg" alt="Image 4" data-target="#demo" data-slide-to="3">
            <img src="img/nightsofcabiria.jpg" alt="Image 5" data-target="#demo" data-slide-to="4">
```

HTML sample, home page.

```
<div class="filter filter-basic">
    <div class="filter-nav">
    <button class="btn btn-success active filter-button" data-filter="">All</button>
    <button class="btn btn-primary filter-button" data-filter="revenue">Box Office Revenue</button>
    <button class="btn btn-primary filter-button" data-filter="budget">Production Budget</button>
    <button class="btn btn-primary filter-button" data-filter="year">Release Year</button>
    <button class="btn btn-primary filter-button" data-filter="genre">Genre</button>
    <button class="btn btn-primary filter-button" data-filter="face">Number of Faces in Poster</button>
    <button class="btn btn-primary filter-button" data-filter="language">Language</button>
    <button class="btn btn-primary filter-button" data-filter="title">Title</button>
    <button class="btn btn-primary filter-button" data-filter="country">Country</button>
    <button class="btn btn-primary filter-button" data-filter="rating">Content Rating</button>
    <button class="btn btn-primary filter-button" data-filter="ratio">Aspect Ratio</button>
    </div>
    <div class="filter-gallery">
    <div class="row">
        <div class="col-md-12 my-1">
        <div class="" data-category="revenue">
            <div class="item-content">…
            </div>
        </div>
        </div>
        <div class="col-md-12 my-1">
            <div class="" data-category="budget">
                <div class="item-content">…
                </div>
            </div>
        </div>
        <div class="col-md-12 my-1">
                <div class="" data-category="year">
                    <div class="item-content">…
                    </div>
                </div>
                </div>
        <div class="col-md-12 my-1">
        <div class="" data-category="genre">
            <div class="item-content">…
            </div>
        </div>
        </div>
        <div class="col-md-12 my-1">
        <div class="" data-category="face">
            <div class="item-content">…
            </div>
        </div>
```

HTML sample, producer's tool page.

```
        <div class="list-group">
            <p class="list-group-item"><b>Year:</b> 2022</p>
            <p class="list-group-item"><b>Runtime:</b> 99 min</p>
            <p class="list-group-item"><b>Genres:</b> Comedy, Drama, Action</p>
            <p class="list-group-item"><b>Production Budget:</b> unknown</p>
            <p class="list-group-item"><b>Box Office Revenue:</b> $32,950</p>
            <p class="list-group-item"><b>Content Rating:</b> Unrated</p>
            <p class="list-group-item"><b>Country:</b> Philippines</p>
            <p class="list-group-item"><b>Language:</b> Tagalog, English</p>
            <p class="list-group-item"><b>Number of Faces in Poster:</b> 1</p>
            <p class="list-group-item"><b>Color:</b> Color</p>
          </div>

        <!--sidebar-->
        <div class="sidebar" id='sidebar'>
            <ul>
                <li><a href="../index.html">Home</a></li>
                <li><a href="../about.html">About</a></li>
                <li><a href="../data.html">Data</a></li>
                <li><a href="../faq.html">FAQ</a></li>
                <li><a href="../contact.html">Contact</a></li>
            </ul>
            <button class="btn btn-primary btn-rounded tool"><a href="../tool.html">Film Success Tool</a></button>
        </div>
    </div>

    <footer class="bg-light py-4">…
    </footer>

    <script>
        var btn = document.querySelector('.toggle');
        var btnst = true;

        btn.onclick = function() {
            if(btnst == true) {
                document.querySelector('.toggle span').classList.add('toggle');
                document.getElementById('sidebar').classList.add('sidebarshow');
                btnst = false;
            } else if(btnst == false) {
                document.querySelector('.toggle span').classList.remove('toggle');
                document.getElementById('sidebar').classList.remove('sidebarshow');
                btnst = true;
            }
        }
    </script>
```

HTML sample, "Leonor Will Never Die" individual data page.

```css
.sidebar ul li a {
  padding:8px 15px;
  font-size:16px;
  color: #222;
  font-family:arial;
  text-decoration:none;
  display:block;
  position:relative;
  z-index:1;
  transition:all 0.3s ease-out;
  font-weight:500;
}

.sidebar ul li a:before {
  content:'';
  position:absolute;
  bottom:0;
  left:50%;
  right:50%;
  transform:translate(-50%,-50%);
  width:0;
  height:1px;
  background: #4CAF50;
  z-index:-1;
  transition:all 0.3s ease-out;
}

.sidebar ul li a:hover:before {
  width:100%;
}

.sidebar ul li a:hover {
  color: #4CAF50;
}

.sidebarshow {
  left:0;
}

.carousel-item {
  width: 98%;
  height: 444px;
  overflow: hidden;
  background-color: #777;
  text-align: center;
```

CSS sample.

Bibliography

Rubin, Rebecca. "Box Office Comeback: Why Movie Theaters Are Finally Feeling Optimistic." Variety, 11 Aug. 2022, variety.com/2022/film/news/box-office-top-gun-elvis-pandemic-return-1235303574/.

Whittington, Richard. "How Film Production Is Becoming More Sustainable and Profitable." Forbes, 2 Mar. 2022, www.forbes.com/sites/sap/2022/02/28/how-film-production-is-becoming-more-sustainable-and-profitable/?sh=704b6edd5bad.

Writers Guild of America West, www.wga.org. 2023.