

Movie Success Prediction Using Naive Bayes, Logistic Regression and Support Vector Machine

Apoorva Shete
Rachaell Nihalani
Prof. Darakhshan Khan

Introduction

The Entertainment Industry

- **The entertainment sector is the leading and most influential sector. Movies have been known to have an extremely strong hold on an individual's mind.**
- **Started as simple medium of entertainment**
- **Quickly grown into a billion-dollar business worldwide**
- **Extremely competitive industry: huge investments both in terms of time and money**
- **The prediction of how a movie will fare at the box office could be a game-changer in the movie sector**
- **Beneficial for shareholders to have access to a model that predicts the success of a movie beforehand, given certain attributes.**
- **Also prove to be highly beneficial to production houses so as to enable them to effectively plan the publicity and promotions, which entails tremendous expenses.**

Movie Success Prediction

- In this paper, we have aimed to predict the success of a motion picture as accurately as possible using a few ML algorithms
- This paper discusses and compares three machine learning algorithms, Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machine (SVM)
- Research objective is to perform analysis and build a model that predicts whether or not a movie would be considered successful, based on certain parameters.
- For efficient evaluation, we train and test our models on two drastically different datasets.
- The performance of these models will be assessed and compared with the help of certain performance measure indices, such as accuracy.
- The results and conclusion of this paper will enable future researchers to select the most effective model, out of the three, that gives the best performance for future applications.

Classifiers

Naive Bayes

- This well-known classification technique is based on Bayes' theorem.
- It assumes:
 - that predictors are independent.
 - that whether or not one feature is present in a class is not at all related to whether another feature is present in it.

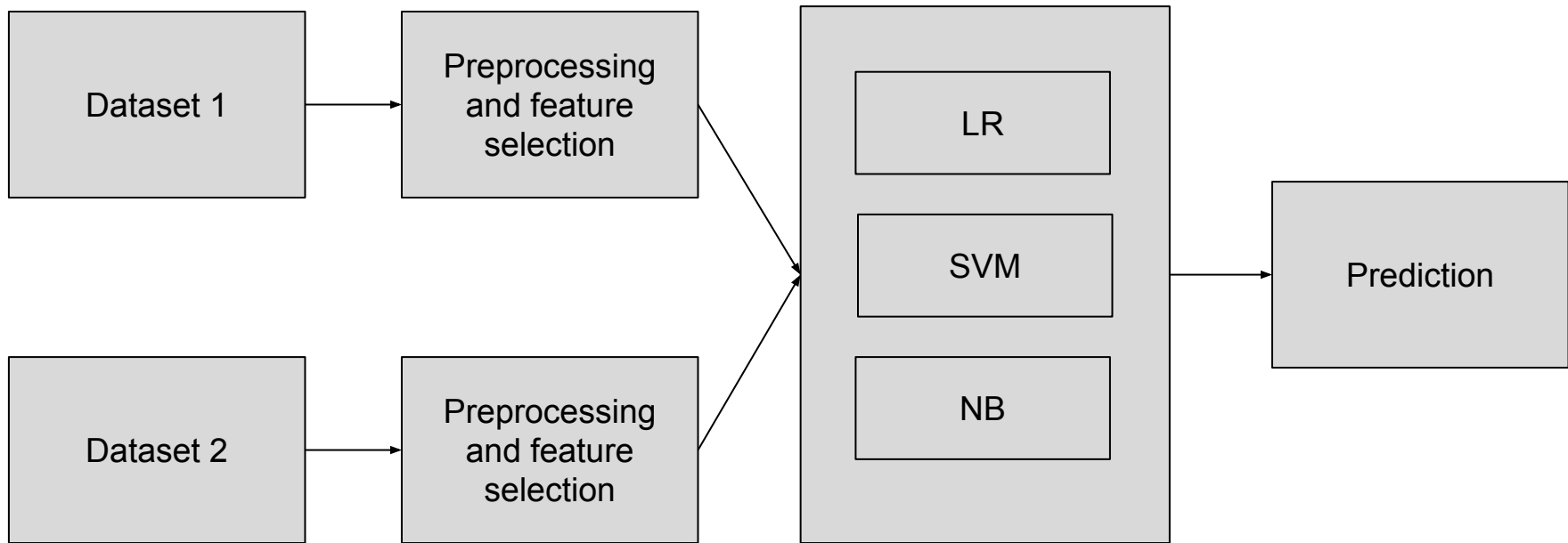
Logistic Regression

- This technique is primarily used for predictive analysis and is built on the probability concept.
- Logistic Regression is a go-to method for binary classification problems.
- It is a linear regression model which does not use a linear function, and instead makes use of the sigmoid or logistic function, which is a cost function of higher complexity.

Support Vector Machine

- SVM is a machine learning algorithm that is used to solve both classifications as well as regression problems. SVM is known for using little computational power to produce considerable accuracy.
- This algorithm aims to find such a hyperplane that distinctly classifies the data points in an N-dimensional space, where N is the number of features.

Methodology



Model Workflow

Results

Our Model has shown

91% and 100%

Highest Accuracy over Dataset 1 and Dataset 2 respectively

Let's see.....

Dataset	Algorithm	Accuracy	Precision	Recall
Dataset 1	Logistic Regression	0.90	0.71	0.72
	Support Vector Machine	0.91	0.79	0.72
	Naive Bayes	0.83	0.52	0.78
Dataset 2	Logistic Regression	1.00	1.00	1.00
	Support Vector Machine	0.99	0.99	1.00
	Naive Bayes	1.00	1.00	1.00

Conclusion

- In today's world where movies lie at the heart of the entertainment society, many young and upcoming producers and artists risk a lot of money when they decide to make an idea into a movie
- Prediction of a movie's success rate before it has been released can be of great help to the people who are investing in making motion pictures
- In our paper, we have successfully compared a few ML algorithms to achieve the best accuracy for predicting a movie's success rate
- It can be seen from the results, the best algorithm for movie success prediction is **Logistic Regression** which has **90%** and **100%** accuracies for **Dataset 1** and **Dataset 2** respectively, giving the **highest accuracy for both the datasets used**. **SVM** has proven to be the **second-best algorithm**, followed by **NB**

Future Scope

- **There is a significant improvement in the success rate of our model as compared to the past research. However, it will give much better accuracy if there is more data used to train this model.**
- **In the future, additional data such as social media comments about the movie's plot can also be utilized for better accuracy.**
- **A movie's success rate can be predicted before it has been released. Once the movie is released, the results of how the movie has fared at the box office can be collected and compared, and analyzed. The results obtained from this analysis can be fed to this ML model to give it better training.**
- **Thus, the prediction rate of this model can be significantly improved in the future.**

References

- [1] N. Quader, M. O. Gani, D. Chaki and M. H. Ali, "A machine learning approach to predict movie box-office success," 2017 20th International Conference of Computer and Information Technology (ICCIT), 2017, pp. 1-7, doi: 10.1109/ICCITECHN.2017.8281839.
- [2] A. Kanitkar, "Bollywood Movie Success Prediction using Machine Learning Algorithms," 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C), 2018, pp. 1-4, doi: 10.1109/CIMCA.2018.8739693.
- [3] R. Dhir and A. Raj, "Movie Success Prediction using Machine Learning Algorithms and their Comparison," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018, pp. 385-390, doi: 10.1109/ICSCCC.2018.8703320.
- [4] J. Ahmad, P. Duraisamy, A. Yousef and B. Buckles, "Movie success prediction using data mining," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2017, pp. 1-4, doi: 10.1109/ICCCNT.2017.8204173.
- [5] N. Darapaneni et al., "Movie Success Prediction Using ML," 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2020, pp. 0869-0874, doi: 10.1109/UEMCON51285.2020.9298145.
- [6] "IMDB-Movie-Data.csv, Dataset 1" <https://github.com/siddrao/Movie-Success-Prediction/blob/master/AIPro/IMDB-Movie-Data.csv>
- [7] "Check.csv, Dataset 2" <https://github.com/SnirZarchi/Movie-Success-Prediction/blob/master/Dataset/check.csv>
- [8] "Logistic Regression." towardsdatascience.com. <https://towardsdatascience.com/introduction-to-logistic-regression-66248243e148>
- [9] "Support Vector Machine" towardsdatascience.com. <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- [10] "Naive Bayes Algorithm" analyticsvidhya.com. <https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/>
- [11] "Classification Outcomes" <https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative>
- [12] "Accuracy" <https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/>
- [13] "Precision and recall." wikipedia.org. https://en.wikipedia.org/wiki/Precision_and_recall

Thank You

Paper Id 564