**Predicting movie success using machine learning techniques**

**ABSTRACT**:

The area of creating predictive models using machine learning has increased in size in recent years. The market for movies is still big with hundreds of new movies created every year. The purpose of this report is to investigate whether it is possible to classify movie rating and box office revenue with metadata available before release. This was done by building a classification model with metadata obtained from the internet such as, budget and what actors are involved, etc. This study managed to correctly predict what rating a movie would have about 70% of the time using the technique with the highest success rate. The results of this report are to a certain extent consistent with previous studies with similar focus in the prediction of the grade. The precision of the predictions can further be increased with a larger data set with more features.

• Classification

• Machine Learning

• Decision Tree

• Support Vector Machine

• KNN

**REFERENCES**:

• Statista Inc. Global box office revenue from 2016 to 2020, 2016. URL https://www. statista.com/statistics/259987/global-box-office-revenue/.

• MathWorks. Classification learner, 2017. URL https://se.mathworks.com/

help/stats/classificationlearner-app.html.

• Roc curves and area under the curve explained, 2017. URL http://www. dataschool.io/roc-curves-and-auc-explained

• Imdb weighted mean rating, 2017. URL <http://www.imdb.com/help/show_>

leaf?votes.z

• Muhammad Hassan Latif and Hammad Afzal. Prediction of movies popularity using machine learning techniques, 2016. <http://paper.ijcsns.org/07_book/>

201608/20160820.pdf