## Related Work:

Data pruning had been defined in 2005 as the automated process of noise cleaning and the performance of this mechanism was measured using SVC and AdaBoost algorithms [1]. Removal of certain portions of the dataset is determined to be worthwhile and said to affect the performance of machine learning algorithms [1].

A mathematical model was proposed to predict the success of upcoming movies based on correlation of factors affecting the success of a movie [2].

Automatic rating prediction was proposed in 2011 using the IMDb dataset, however the results were inferior to baseline which was attributed to the dataset lacking diversity in terms of user rating [3].

## References

|  |  |
| --- | --- |
| [1] | A. Angelova, Y. Abu-Mostafa and P. Perona, "Pruning training sets for learning of object categories," in *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005. |
| [2] | J. Ahmad, P. Duraisamy, A. Yousef and B. Buckles, "Movie success prediction using data mining," in *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2017. |
| [3] | P. Biljanović and E. a. M.-M. Croatian Society for Information and Communication Technology, MIPRO 2011 : 34th International Convention on Information and Communication Technology, Electronics and Microelectronics : May 23-27, 2011, Opatija, Croatia : proceedings, Croatian Society for Information and Communication Technology, Electronics and Microelectronics, 2011. |
| [4] | V. Subramaniyaswamy, M. V. Vaibhav, R. V. Prasad and R. Logesh, "Predicting movie box office success using multiple regression and SVM," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 2017. |
| [5] | P.-Y. Hsu, Y.-H. Shen and X.-A. Xie, "LNCS 8818 - Predicting Movies User Ratings with Imdb Attributes," *LNAI,* vol. 8818, pp. 444-453, 2014. |
| [6] | "How to One Hot Encode Sequence Data in Python," [Online]. Available: https://machinelearningmastery.com/how-to-one-hot-encode-sequence-data-in-python/. |
| [7] | "Data Cleaning 101 – Towards Data Science," [Online]. Available: https://towardsdatascience.com/data-cleaning-101-948d22a92e4. |