Gazi AKADEMİK YAYINCILIK

ARTIFICIAL INTELLIGENCE STUDIES

# Prediction With Deep Learning Neural Networks: The Careers in Show Business

**Ali Can Günhan** [*a], **Kamil Topal** [b]

## ABSTRACT

In this article, we investigate whether we can predict individual success in the film industry by using four distinct deep learning neural networks. It is shown that the highest rates of accuracy in prediction can be obtained through using the bidirectional deep learning algorithms. We found that when the prediction is taken into account, there is no gender bias. These findings can be explained by the fact that the film industry is essentially dominated by the popularity for both actors and actresses. Moreover, since popularity, to a greater extent, determines success, bidirectional algorithms are more effective in predicting success due to the fact that they are able to take into account both past and future information regarding a particular data point. This is a must in predicting success in the film industry, since popularity and its lack thereof determines success and failure in the past as well as in the future of an acting career.

a,* Mersin University,
Faculty of Science,
Dept. of pyhsics
33343 - Mersin, Türkiye
ORCID: 0000-0003-0050-2484

a,* Balıkesir University,
Faculty of Engineering,
Dept. of Computer Engineering
10145 - Balıkesir, Türkiye
ORCID: 0000-0002-0266-7365

* Corresponding author.
e-mail: alicangunhan@mersin.edu.tr

**Keywords:** Actor/Actress, Film Industry, Deep Learning, Prediction Success,

# Derin Öğrenme Sinir Ağları ile Tahmin: Gösteri Dünyasında Kariyer

## ÖZ

Bu makalede, dört farklı derin öğrenme sinir ağı kullanarak film endüstrisindeki bireysel başarıyı tahmin edip edemeyeceğimizi araştırıyoruz. Tahminde en yüksek doğruluk oranlarının çift yönlü derin öğrenme algoritmaları kullanılarak elde edilebileceği gösterilmiştir. Tahmin dikkate alındığında cinsiyet yanlılığının olmadığını bulduk. Bu bulgular, film endüstrisinin esasen hem aktörler hem de aktrisler için gözde oluşun hâkim olduğu gerçeğiyle açıklanabilir. Ayrıca, gözde oluş (popülerlik) büyük ölçüde başarıyı belirlediğinden, çift yönlü algoritmalar, belirli bir veri noktasıyla ilgili hem geçmiş hem de gelecekteki bilgileri hesaba katabildikleri için başarıyı tahmin etmede daha etkilidirler. Bu, film endüstrisindeki başarıyı tahmin etmede bir zorunluluktur, çünkü gözde oluş ve bunun eksikliği oyunculuk kariyerinin hem geçmişteki hem de geleceğindeki başarıyı ve başarısızlığı belirler.

**Anahtar Kelimeler:** Aktör/Aktris, Film Sanayii, Derin Öğrenme, Tahmin, Başarı

Submitted: 16.07.2023
Revised: 04.10.2023
Accepted: 24.10.2023

doi: 10.30855/gmbd.2023.06.02.01

**To cite this article**: **To cite this article**: Ali Can Günhan, Kamil Topal, "Prediction With Deep Learning Neural Networks:The Careers in Show Business," *Artificial Intelligence Studies,* vol.6, no.2, pp. 14-20, 2023. doi: 10.30855/gmbd.2023.06.02.01

## 1. Introduction (Giriş)

In recent years, we have seen new ways to apply computational and mathematical tools to social phenomena. Some of those phenomena are information spreading in social networks [1], agent-based models for interactions between individuals and collective items [2], and terrorism [3]. To analyze and quantify success -and equivalently failure- within society [4, 5, 6, 7, 8] is an example for application of computational based studies in social sciences. Machine learning and deep learning methods have been effective in such analyses [9, 10, 11]. This methodology has allowed us to accurately foresee the evolution of social phenomena that would otherwise not be easy to predict.

In references [12] and [13] the success of actors and actresses has been predicted with high accuracy using machine learning algorithms. In [12] the worldwide International Movie Database (IMDb) database is studied while in [13] authors considered real data in the Turkish movie industry. This made us to study same phenomena with deep learning algorithms.

In this work, we studied the success of individual actresses and actors in the movie industry. We quantified and predicted success of acting careers. The goal of our prediction is to determine the peak in the career of the actors and actresses, called *annus mirabilis* (miracle year). For this purpose, we made use of data from IMDb. Using 4 distinct deep learning (DL) algorithms, we have predicted annus mirabilis (AM) of performers.

The paper is organized as follows. In the next section, we provide the method and briefly explain the DL algorithms used throughout the research. In section 3, we present the results obtained from the respective DL algorithms. The conclusions are presented in section 4.

## 2. Method and Algorithms (Yöntem ve Algoritmalar)

As noted above, the dataset is from IMDb [14]. The actors and actresses whose active careers exceed 20 years are selected from the IMDb dataset. Then, we have determined the threshold value to be a minimum 5 credits so that the *annus mirabilis* (AM) corresponds to the year with 5 or more credits. Five variables for each actor or actress sampled from their career length are randomly created. These variables are used to form sub-careers whose initial point is the same as the beginning of the actual career under scrutiny. Then, the end values of the sub-careers are randomly fixed based on the five variables of the aforementioned actor and actress. For each sub-career, a particular set is tagged as 1 or 0 depending on whether AM is included in that particular set or not, respectively. This outcome is randomly generated 100 times and 5-fold cross validation has been carried out. The scores useful to quantify the data read,

$$Precision \equiv \frac{TP}{TP+FP} \ , \tag{1}$$

$$Recall \equiv \frac{TP}{TP+FN} \ , \tag{2}$$

$$Accuracy \equiv \frac{TP+TN}{TP+FP+TN+FN}, \tag{3}$$

where True/False Positives and True/False Negatives are denoted as TP/FP and TN/FN, respectively. Note that F-score is defined as the harmonic average of Precision and Recall,

$$F - score \equiv \frac{2TP}{2TP+FP+FN}. \tag{4}$$

Using the final dataset as the training the data, we apply the following DL algorithms to predict the AM. The first DL algorithm is the Gated Recurrent Unit (GRU) whose objective is to resolve the vanishing gradient problem in standard recurrent neural networks (RNN) through update gate and reset gate [15]. The former (update) gate filters the past information from the previous time steps irrelevant to prediction. The latter (reset) gate enables to forget the unfiltered past information. In this way algorithm chooses which information should be passed to the output. The second DL algorithm is the Bidirectional Gated Recurrent Unit (biGRU). It is a GRU algorithm that possesses additionally a future layer so that it can extract information concerning both past and future [16]. The next DL algorithm is the Long Short-Term Memory (LSTM) recurrent neural networks which are devised for situations that require long term temporal dependence [17]. Finally, the Bidirectional Long Short-Term memory (biLSTM) provides every point in the sequence with information before and after it [18].

In applying the algorithms, we used TensorFlow library in Python language. Each of the algorithms has an input layer, 50 units of hidden layers, 10 units of hidden layers and 1 unit of output layer. These models were tested with learning coefficients of 0.001 and 0.0001. Sigmoid activation function, Adam optimization, and for the loss function transverse entropy are used.

## 3. Predicting Annus Mirabilis with Deep Learning Algorithms (Derin Öğrenme Algoritmaları ile Mucize Yılı (*Annus Mirabilis*) Tahmin Etme)

For comparison, we provide the results in Reference [12] which were obtained through machine learning, denoted as ML. All the precise results are tabulated in Table 1.

Tablo 1. Precise results that obtained in applied algorithms. (Uygulanan algoritmalarda elde edilen kesin sonuçlar.)

| | ML | | LSTM | | GRU | | biLSTM | | biGRU | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female |
| *Precision* | 0.86 | 0.83 | 0.58 | 0.74 | 0.58 | 0.65 | 0.85 | 0.87 | 0.86 | 0.89 |
| *Recall* | 0.76 | 0.78 | 0.99 | 0.93 | 0.99 | 0.99 | 0.97 | 0.98 | 0.97 | 0.98 |
| *Accuracy* | 0.84 | 0.86 | 0.58 | 0.73 | 0.58 | 0.65 | 0.89 | 0.90 | 0.90 | 0.92 |
| *F-scores* | 0.81 | 0.80 | 0.73 | 0.82 | 0.73 | 0.79 | 0.90 | 0.92 | 0.91 | 0.93 |

The results of our prediction concerning the show business careers through four distinct deep learning algorithms are given in Figures 1-4. These figures show the precision, recall, accuracy and F-scores, respectively. As can be seen from these figures, LSTM and GRU algorithms provide close or lower values to the results in Reference [12] for the precision, recall, accuracy and F-scores. However, when bidirectional DL algorithms are implemented, this behavior changes to the benefit of the DL learning algorithms. In particular, we note that the most important quantitative measure for prediction is the accuracy score given in Fig. 3, since it is the one which accords the greater values of TP and TN. From Fig. 3 then, we see that biGRU yields the highest values.

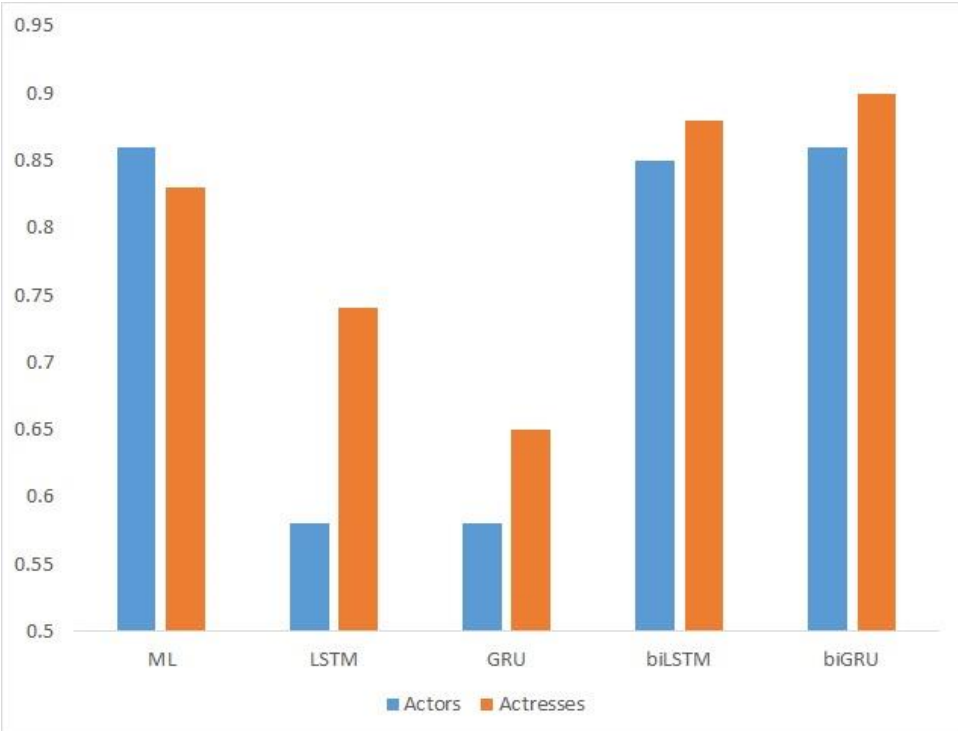


Figure 1. The precision scores that obtained with machine learning (ML) and four distinct deep learning (DL) algorithms. (Makine öğrenimi (ML) ve dört farklı derin öğrenme (DL) algoritması ile elde edilen hassasiyet puanları.)

The precision score is used to decide whether we have low ratios of FP. In particular, the precision values for the actors are almost same whether one adopts ML or bidirectional DL approach as can be seen from Fig. 1. However, LSTM and GRU drastically fails in precision. From Fig. 2, one can observe that all DL algorithms are yielding better outcomes of recall scores compared to the ML approach. As a matter of fact, even though it is lack of bidirectional feature, GRU functions as good as biLSTM and biGRU in prediction success. The recall score is useful in assessing whether one has low FN or not. In comparing the F-scores in Fig. 4, we also observe here that bidirectional DL algorithms are more successful. This is to be expected, however, since F-scores are obtained by evaluating the harmonic average of the precision and recall scores, and in both scores, bidirectional DL algorithms provide better values as can be seen from Figs. 1 and 2.

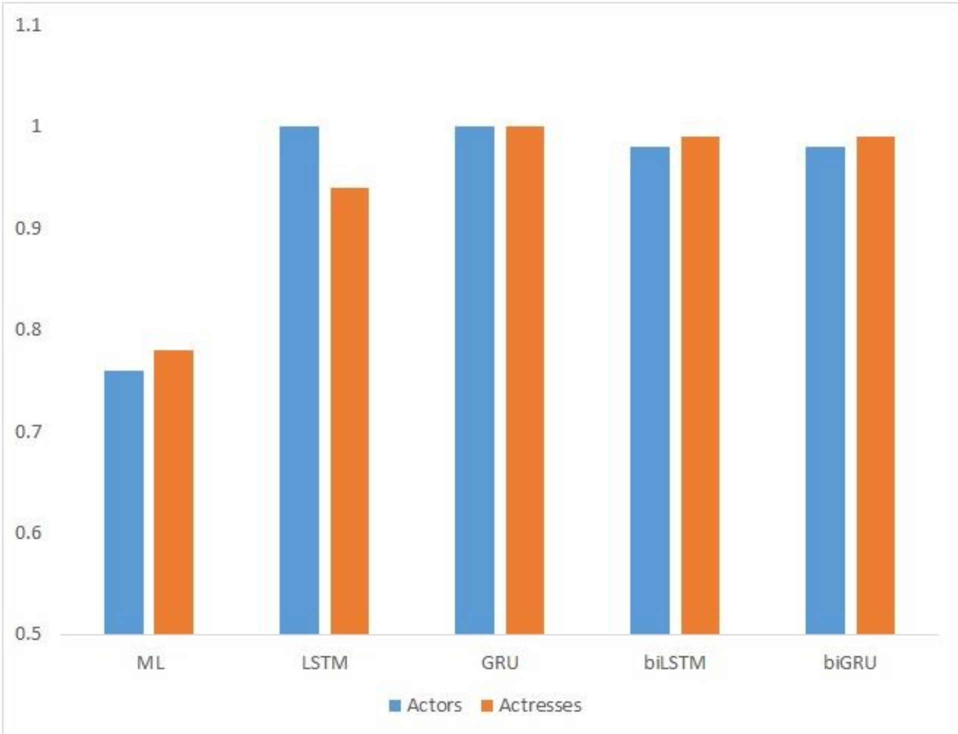


Figure 2. The recall scores that obtained with machine learning (ML) and four distinct deep learning (DL) algorithms.
(Makine öğrenimi (ML) ve dört farklı derin öğrenme (DL) algoritması ile elde edilen hatırlama puanları.)

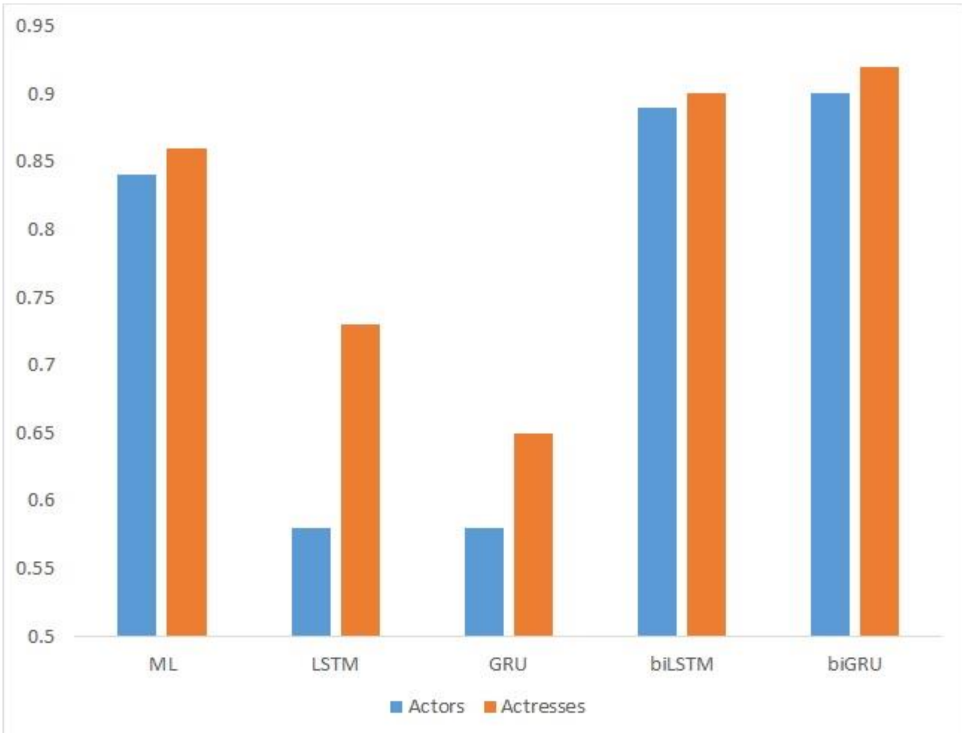


Figure 3. The accuracy scores that obtained with machine learning (ML) and four distinct deep learning (DL) algorithms.
(Makine öğrenimi (ML) ve dört farklı derin öğrenme (DL) algoritması ile elde edilen kesinlik puanları.)

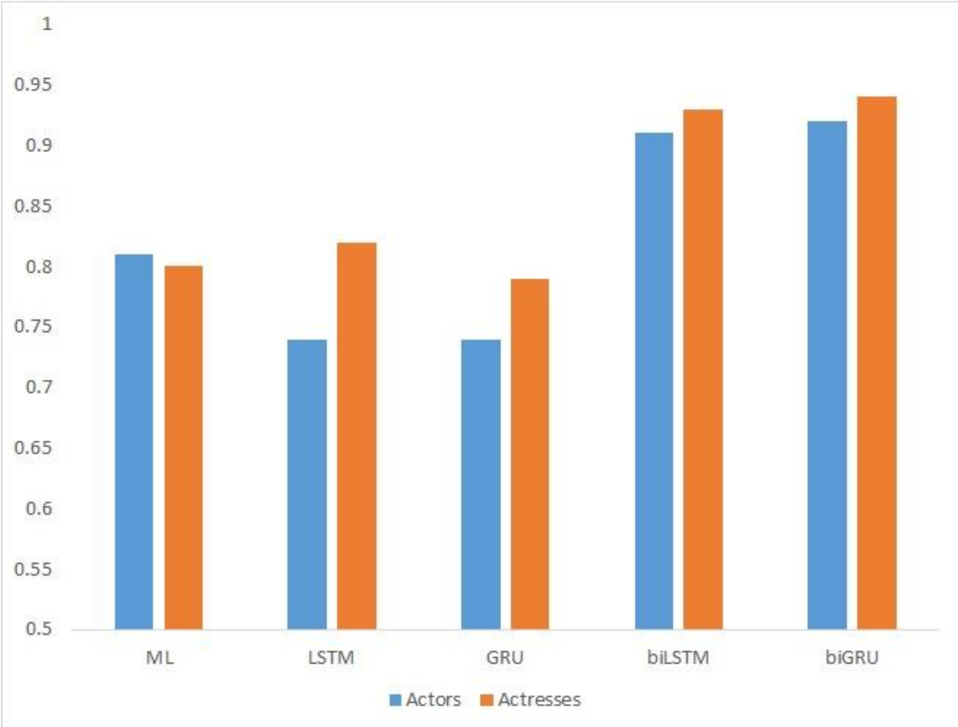


Figure 4. The F-scores that obtained with machine learning (ML) and four distinct deep learning (DL) algorithms.
(Makine öğrenimi (ML) ve dört farklı derin öğrenme (DL) algoritması ile elde edilen F-skor puanları.)

Although show business data is shown to be biased in gender as in [12, 13], we see no sign of an asymmetry in bidirectional DL results observed in Figs. 1-4. However, LSTM and GRU yield very different precision and accuracy scores for actors and actresses. The careers of the actresses have better accuracy scores compared to actors.

## 4. Results and Discussion (Sonuçlar ve Tartışma)

In this paper, we studied the individual success in show business. Using four distinct deep learning algorithms, the year with the maximum number of show credits in the career of an actor or actress (i.e., *annus mirabilis*) is predicted. These algorithms are the Gated Recurrent Unit (GRU), the Bidirectional Gated Recurrent Unit (biGRU), the Long Short-Term Memory (LSTM), the Bidirectional Long Short-Term memory (biLSTM). The highest accuracy scores are found to be higher than the machine learning results given in Ref. [12]. Moreover, the highest accuracy is obtained when one uses the bidirectional algorithms, i.e., biLSTM gives 0.89 accuracy for male, 0.90 accuracy for female perfomers, and biGRU gives 0.90 accuracy for males and 0.92 accuracy for females. This is reasonable, since the success in the show business highly depends on the popularity and this requires the analysis of the data points in a bidirectional manner. Indeed, popularity is important to both success and failure in show business, and so we can make more accurate predictions if data points are analyzed backwards as well as forwards.

Although it has already been observed that show business data exhibits gender difference [12, 13], we have not observed such an asymmetry in the accuracy scores considering the bidirectional deep learning algorithms. This implies that gender difference plays no role as far as one focuses on the prediction of the success of the individuals. This observation in our work can be explained by noting that the essential element for success in show business, namely popularity, applies equally well to both actors and actresses.

Finally, through this paper, we emphasize that the notion of success is predictable and possesses an underlying pattern. This implies that one can rely on deep learning algorithms to foresee a possible turnover in the career of an individual and make recommendations accordingly. Also, it is worth noting that since predicting success and failure is possible, it must also be possible to mathematically model them [19].

## Acknowledgment (Teşekkür)

It is a pleasure to thank G. Barış Bağcı for exciting our interest in network science and for helpful discussions. One of us, ACG, thanks to Arinna B. Günhan for her help in browsing the web crawler and İlayda Ağar and Robert J. Gonzalez for editing.

## Conflict of Interest Statement (Çıkar Çatışması Beyanı)

The authors declare that there are no relevant financial or non-financial competing interests.

## References (Kaynaklar)

[1] J. L. Iribarren and E. Moro, "Branching dynamics of viral information spreading," *Phys. Rev. E*, vol. 84, p. 046116, Oct 2011. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.84.046116

[2] J. M. Epstein, "Agent-based computational models and generative social science," *Complexity*, vol. 4, no. 5, pp. 41–60, 1999. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291099-0526%28199905/06%294%3A5%3C41%3A%3AAID-CPLX9%3E3.0.CO%3B2-F

[3] A.T. Turk, "Sociology of terrorism", *Annual Review of Sociology*, vol.30, no.1, pp. 271–286, 2004.

[4] M. Janosov, F. Battiston and R. Sinatra, "Success and luck in creative careers," *EPJ Data Science*, vol. 9, no. 1, pp. 9–20, 2020.

[5] A. Pluchino, A. E. Biondo, and A. Rapisarda, "Talent versus luck: the role of randomness in success and failure," *Advances in Complex Systems*, vol. 21, no. 03n04, p. 1850014, 2018.

[6] A. Pluchino, G. Burgio, A. Rapisarda, A. E. Biondo, A. Pulvirenti, A. Ferro and T. Giorgino, "Exploring the role of interdisciplinarity in physics: Success, talent and luck," *PLoS ONE*, vol. 14, no. 6, p. e0218793, 2019.

[7] P.C. Petrantonakis and I. Kompatsiaris, "On the talent vs. luck-based evaluation of the classification process," *IEEE Access*, vol. 7, pp. 37 565–37 574, 2019.

[8] P. Sobkowicz, R. H. Frank, A. E. Biondo, A. Pluchino and A. Rapisarda, "Inequalities, chance and success in sport competitions: Simulations vs empirical data," *Physica A: Statistical Mechanics and its Applications*, vol. 557, p. 124899, 2020.

[9] R. Sharda and D. Delen, "Predicting box-office success of motion pictures with neural networks," *Expert Systems with Applications*, vol. 30, no. 2, pp. 243–254, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417405001399

[10] O. Kart, O. Ulucay, B. Bingol and Z. Isik, "A machine learning-based recommendation model for bipartite networks," *Physica A: Statistical Mechanics and its Applications*, vol. 553, p. 124287, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437120300844

[11] L. Zhang, J. Luo and S. Yang, "Forecasting box office revenue of movies with bp neural network," *Expert Systems with Applications*, vol. 36, no. 3, Part 2, pp. 6580–6587, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095741740800496X

[12] O. E. Williams, L. Lacasa and V. Latora, "Quantifying and predicting success in show business," *Nature Communications*, vol. 10, no. 1, p. 2256, 2019. [Online]. Available: https://doi.org/10.1038/s41467-019-10213-0

[13] K. Topal, A. C. Günhan and G. B. Bagci, "Predicting *annus mirabilis* with machine learning: Turkish movie industry," *Multimedia Tools and Applications*, 2023. [Online]. Available: https://doi.org/10.1007/s11042-023-16212-0

[14] International Movie Database (IMDb). Date accessed: 2020.05.26 https://developer.imdb.com/non-commercial-datasets

[15] K. Cho, B. van Merriënboer, Ç. Gülçehre, F. Bougares, H. Schwenk and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, A. Moschitti, B. Pang, and W. Daelemans, Eds. Doha, Qatar: Association for Computational Linguistics, oct 2014, pp. 1724–1734. [Online]. Available: https://aclanthology.org/D14-1179/

[16] F. Liu, J. Zheng, L. Zheng and C. Chen, "Combining attention-based bidirectional gated recurrent neural network and two-dimensional convolutional neural network for document-level sentiment classification," *Neurocomputing*, vol. 371, pp. 39–50, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092523121931272X

[17] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 11 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[18] M. Gao, G. Shi and S. Li, "Online prediction of ship behavior with automatic identification system sensor data using

bidirectional long short-term memory recurrent neural network," *Sensors*, vol. 18, no. 12, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/12/4211

[19] Y. Yin, Y. Wang, J. A. Evans and D. Wang, "Quantifying dynamics of failure across science, startups, and security," *Nature*, vol. 575, no. 7781, pp. 190–194, 2019.