**NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS**

**FACULTY**

**COMPUTER SCIENCE**

**PROJECT PROPOSAL**

**DEVELOPMENT OF MVP MACHINE LEARNING SYSTEM FOR PREDICTING THE PROFITABILITY OF THE FILM BUSINESS**

Elizar Chepokov, SE-18-2

**Supervisor:** Professor of the Department of

Information Technology in Business,

Doctor of Technical Sciences,

Leonid Yasnitsky

**PERM**

**2022**

Currently, the film business is not going through the best of times, because of the current epidemiological situation, films around the world are postponed or completely switched to streaming services, which certainly affects the box office of films. This study is aimed at studying the causal relationships between the success of a film and a variety of variables, for a scientifically based assessment of the profitability of the film business. This paper describes the analysis of previous research in this field and existing solutions, as well as the technology stack and the architecture design of the system. This development uses the methods of software construction and UML design. The result of this development is a system using a trained neural network, with the ability to enter input data by the user.

# **Introduction**

**Background.** The film industry makes a huge contribution both to the general social life of a person and to the development of the economy of countries. According to the report of the Motion Picture Association of the American Association of Film Companies [7], in 2020, due to the epidemiological turmoil that affected the entire world, box office receipts fell sharply and amounted to $80.8 billion, 17.8% less than in 2019 ($98.3 billion), the number of tickets sold was 223 million, against 1.3 billion in 2019. It is also worth noting that the share of cinema rentals is decreasing every year, when the share of online viewing on streaming services is growing, so, in 2020, streaming services collected $61.8 billion (76% of total box office receipts), which is $14.6 billion more than in 2019 ($47.2 billion). Film production has been and remains one of the riskiest types of entrepreneurship, especially in times of restrictions on visiting public places. Spending huge budgets, film companies expect corresponding revenues from showing the film in cinemas and sales on streaming services, but investors note that it is extremely difficult to predict the financial success of the film even for experienced market participants. The relevance of this final qualifying work lies in the fact that the release of a motion picture nowadays is a big risk, not only for film studios and directors, but also for investors, the forecasting system is able to reduce the risks of unsuccessful investments in film production at the early stages of the creation of a motion picture, as well as predict the feasibility of releasing a motion picture on cinema screens with the associated costs of advertising and rental of cinemas.

**Problem Statement.** Currently, decisions on the feasibility of implementing film projects are made mainly on the basis of intuition, ignoring user criticism and focusing on the success of the previous franchise, which contributes to an unjustified increase in the risks of investors investing significant financial resources in film production. When, in order to increase the validity of the decisions taken [4], it is necessary that they be based on a comprehensive analysis of the strengths and weaknesses of the film project before it is put into production. The research is based on the need for scientific substantiation and creation of an effectively functioning system for forecasting the commercial potential of a film project, determining the directions and principles of effective management of the cinematic business process, in particular, at the early stages of making films.

**Delimitations of the Study.** The object of research in this paper is the machine learning algorithm system. The subject of the research is the development of a machine learning system algorithm. The purpose of this work is to develop a machine learning system for predicting the profitability of the film business. The implementation requires to perform the following tasks:

1. To analyze the area of interest, which includes analyzing the film business and options for sorting films by parameters, analyzing the possibilities of building neural networks and integrating them with the system, formalizing existing methods for predicting the profitability of the film business and specifying the requirements for the system being developed.
2. Design a system by designing a data collection and systematization module and designing a neural network module with algorithms for selecting the importance of variables and finding the profitability of the film business.
3. Implement a machine learning system to predict the profitability of the film business.

**Professional Significance.** The practical significance of this work lies in the possibility of using the system being developed to present methodological recommendations for the rationalization of production activities that ensure the achievement of high economic results of the film industry and individual films.

The scientific novelty of the work consists in the development of models, approaches and techniques, as well as a system that contributes to the analysis of the commercial success of films, allowing us to draw quantitatively sound conclusions about their commercial potential.

# **Literature Review**

These days, a rapid growth of computing technology and the emergence of new technologies provides for new ways to generate data as well as in some cases requiring new methods to process this data when traditional means are insufficient or unfeasible. Some of the first researchers who applied methods of economic and mathematical modeling in the film business were J.Prag and J.Casavant [9], who in 1994 published an article reporting on the creation of a regression model based on a sample of 625 American films. The set of input variables they considered included production budget, critical reviews, availability of stars, franchise availability, availability of awards, genre and rating.

In 2002 American scientists R.Sharda and D.Delen [10] developed the same study using neural networks for predicting the box office of films. In 2006, the same authors built models based on logistic regression, discriminant analysis, classification and regression trees and also converted a predictive problem into a classification problem. In 2009 [11], they also decided to improve forecasting by including decision trees and a more comprehensive sampling of movies.

In 2009 W. Chang and K.J. Lee [1] proposed using a Bayesian belief network to identify causal relationships in predicting the box office success of Korean films. The number of viewers subdivided into groups was chosen as an indicator of success. Compared to neural networks and decision trees, this approach turned out to be more accurate.

Further research has also been conducted into finding new cause-and-effect relationships between the box office success of films and a variety of variables. Variants of the relationship between the activity of users in social media such as Twitter [12] and YouTube [13] on the rating and popularity of the film were proposed by researchers A. Oghina, M. Breuss et al. [8] in 2012. Furthermore, researchers conducted a sentiment analysis of Twitter users' comments and their causal relationship on the rating of a film on the IMDb [3] platform [2].

In the papers described above, little emphasis is placed on the application of the totality of all accumulated knowledge, as well as insufficient importance to social media and advertising. In this paper, the sample for prediction will be expanded, along with the use of technologies and correlations used in the described works as well as new relationships for profitability of the film business extracted from the variables will also be put forward.

# **Methods**

The development of a system for forecasting the profitability of films, it is necessary to start with an analysis of the film business industry, as well as with existing solutions and their components for implementing the functionality of the program and forming the technological stack of the system [5]. Together with the analysis, the following research methods will be applied: synthesis, systematic approach, statistical analysis, comparison methods.

After that, modeling methods should be used to create and formalize the application architecture. These methods contain various approaches and designations to describe software development. The database structure and tables should be visualized using an entity relationship diagram to cover all the objects that will be stored in the database and determine their relationships. The architecture design must conform to the UML standard for creating diagrams, which should include: use case diagram to represent user interaction with the system, class diagram to describe the structure of the system, which will divide the functionality of the application into logical parts and explain their interrelationships and sequence diagram showing interactions within the system over time, revealing the actions and order of interaction between the components of the system.

In the process of implementing the application, the following methods of building software will be used: minimizing complexity, anticipating changes, building for verification and reuse. In addition, the code will conform to the concepts of object-oriented programming (OOP) [6]. The developed application should use unit testing to verify that each part of the program is working properly. Integration testing will be applied to find errors in the interaction of components.

# **Results Anticipated**

This development is aimed at creating a neural network with a user interface for predicting the profitability of a particular movie. When modeling and developing the system, the achievements of researchers in this field will be used, as well as the selection will be expanded and new relationships between variables will be added. In addition, the research paper will contain an analysis of the research area, including a list of relevant technologies that can be used to develop the system.

The results will include all development artifacts, such as system requirements and technology stack, architecture design schemes, testing documentation, deployment, use and maintenance of the system. The architecture design will be documented in the UML standard, will contain schemas of use cases, classes and sequences. The main part of the expected results is a stably working trained neural network that will produce an estimated result based on the data entered by the user.

# **Conclusion**

The implemented neural network can be used for commercial use by directors, companies and investors to predict upcoming movies. So there is an opportunity to modernize the system, with the ability to output information that will increase the popularity of the film. Also it is possible that this study will help in future research of neural networks, for a fully modernized logic of neural networks.

# **References**

Breuss, M., de Rijke, M., Oghina, A., & Tsagkias, M., 2012.

“Prediciting IMDB movie ratings with social media”. In proceedings of

the 34th European Conference on IR Research: Advances in Information

Retrieval, ECIR, Barcelona, Spain, Eds. R. Baeza-Yates & A.P. de Vries

& H. Zaragoza & B.B. Cambazoglu & V. Murdock & R. Lempel & F.

Silvestri, LNCS vol 7224, pp. 503-507, Springer.

Breuss, M., de Rijke, M., Oghina, A., & Tsagkias, M., 2012.

“Prediciting IMDB movie ratings with social media”. In proceedings of

the 34th European Conference on IR Research: Advances in Information

Retrieval, ECIR, Barcelona, Spain, Eds. R. Baeza-Yates & A.P. de Vries

& H. Zaragoza & B.B. Cambazoglu & V. Murdock & R. Lempel & F.

Silvestri, LNCS vol 7224, pp. 503-507, Springer.

Breuss, M., de Rijke, M., Oghina, A., & Tsagkias, M., 2012.

“Prediciting IMDB movie ratings with social media”. In proceedings of

the 34th European Conference on IR Research: Advances in Information

Retrieval, ECIR, Barcelona, Spain, Eds. R. Baeza-Yates & A.P. de Vries

& H. Zaragoza & B.B. Cambazoglu & V. Murdock & R. Lempel & F.

Silvestri, LNCS vol 7224, pp. 503-507, Springer.

Breuss, M., de Rijke, M., Oghina, A., & Tsagkias, M., 2012.

“Prediciting IMDB movie ratings with social media”. In proceedings of

the 34th European Conference on IR Research: Advances in Information

Retrieval, ECIR, Barcelona, Spain, Eds. R. Baeza-Yates & A.P. de Vries

& H. Zaragoza & B.B. Cambazoglu & V. Murdock & R. Lempel & F.

Silvestri, LNCS vol 7224, pp. 503-507, Springer.

1. W. Chang and K.J. Lee, “Bayesian belief network for box-office performance: A case study on Korean movies,” in *Expert Systems with Applications*, vol. 36(1), Elsevier, 2009, pp. 280–291.
2. M. Ghiassi, J. Skinner and D. Zimbra, “Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network,” in *Expert Systems with Applications*, vol. 40(16), 2013, pp. 6266–6282.
3. Internet Movie Database, 2022. Retrieved from <https://www.imdb.com/>.
4. A. Kaur and A.P. Nidhi, 2013. “Predicting movie success using Neural Network” in *International Journal of Science and Search*, vol. 2(9), India, 2013, pp. 69-71.
5. Z. Li, L. Jianhua and Y. Su ying, 2009. “Forecasting box office revenues of movies with BP neural network” in *Expert Systems with Applications*, vol. 36(3), Elsevier, Part 2, 2009, pp. 6580–6587.
6. S. McConnell, “Code complete,” Redmond, WA: Microsoft Press, 2016.
7. Motion Picture Association “THEME Report 2020,” 2021.
8. A. Oghina, M. Breuss, M. Tsagkias and M. de Rijke, “Prediciting IMDB movie ratings with social media,” in *34th European Conference on IR Research* *ECIR 2012: Advances in Information Retrieval*, 2012, pp. 503-507, doi: 10.1007/978-3-642-28997-2\_51.
9. J. Prag and J. Casavant, “An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry,” in *Journal of Cultural Economics*, vol. 18(3), 1994, pp. 217–235.
10. R. Sharda and D. Delen, “Predicting box-office success of motion pictures with neural networks,” in *Expert Systems with Applications*, vol. 30, 2006, pp. 243–254
11. R. Sharda and D. Delen, “Predicting the financial success of Hollywood movies using an information fusion approach,” in *Industrial Engeneering Journal*, vol. 21, 2010, pp. 30–38.
12. Twitter, 2022. Retrieved from <https://twitter.com/>
13. YouTube, 2022. Retrieved from <https://www.youtube.com/>