

УДК 791.43

DOI: 10.30628/1994-9529-2024-20.2-55-84

EDN: LXOADM

Статья получена 09.04.2024, отредактирована 12.06.2024, принята 28.06.2024

**АНТОН ВАЛЕНТИНОВИЧ ДОЖДИКОВ**

Институт социально-политических исследований  
Федерального научно-исследовательского  
социологического центра Российской академии наук  
Россия, 11933, г. Москва, ул. Фотиевой, д. 6, к. 1

**ResearcherID: KYP-9166-2024**

**ORCID: 0000-0002-1069-1648**

**e-mail: antondnn@yandex.ru**

*Для цитирования*

Дождиков А.В. Повышение эффективности государственной политики в сфере кинематографа с помощью машинного обучения // Наука телевидения. 2024. 20 (2). С. 55–84. DOI: 10.30628/1994-9529-2024-20.2-55-84. EDN: LXOADM

## Повышение эффективности государственной политики в сфере кинематографа с помощью машинного обучения

**Аннотация.** Проведено исследование массива прокатных данных российских национальных кинофильмов с февраля 2004 по сентябрь 2023 года с применением методов машинного обучения: отдельно рассмотрены успешные и неуспешные в прокате фильмы, проекты патриотической направленности. В исследовании использована ансамблевая модель машинного обучения HistGradientBoostingClassifier и последовательная полносвязная трехслойная нейросеть на основе библиотеки TensorFlow, базовые методы обработки естественного языка. Установлено, что патриотические фильмы имеют прокатные характеристики ниже, чем в среднем по рынку, и существенно отстают от успешных в прокате проектов российского кинематографа. Доказана возможность точного прогнозирования киносборов, прокатных характеристик,

© Наука  
телевидения



а также подбора параметров проекта и состава его творческой группы для улучшения результатов проката и увеличения охвата целевых аудиторий. Показано направление использования языковых моделей (на примере аннотаций кинофильмов) для создания эффективного киноконтента. Сделан вывод о необходимости комплекса мер по реализации государственной политики в сфере образования, информационных технологий, культуры и искусства, которая должна быть основана на создании единых медиафраншиз и «зонтичных брендов», включающих кинофильмы, сериалы, компьютерные, онлайн и настольные игры, музыкальную и иную продукцию. Внедрение результатов исследования повысит эффективность государственной политики, отдачу от государственных и частно-государственных инвестиций в сфере креативных индустрий. Кинематограф и креативная индустрия станут самоокупаемыми и приносящими доход секторами экономики. Использование данных проката других рынков позволит создавать более востребованный за рубежом контент как элемент «мягкой силы» и культурного влияния. Результаты исследования могут использоваться во время отбора и разработки кинопроектов «Фондом кино», Министерством культуры Российской Федерации, частными российскими инвесторами и киностудиями, органами власти и организациями, ответственными за распространение российского киноконтента за рубежом.

**Ключевые слова:** государственная политика, национальный кинематограф, обработка естественного языка, языковая модель, машинное обучение, искусственный интеллект, нейросеть, «мягкая сила», политика в сфере культуры, Фонд кино, Министерство культуры Российской Федерации, Россотрудничество, Росконгресс

---

**UDC 791.43**

DOI: 10.30628/1994-9529-2024-20.2-55–84

EDN: LXOADM

Received 30.01.2024, revised 18.03.2024, accepted 28.06.2024

**ANTON V. DOZHDIKOV**

Institute of Socio-Political Research,  
Federal Center of Theoretical and Applied Sociology  
of the Russian Academy of Sciences,  
6, korp. 1, Fotiyevoy, Moscow 11933, Russia

**ResearcherID: KYP-9166-2024**

**ORCID: 0000-0002-1069-1648**

**e-mail: antondnn@yandex.ru**

*For citation*

Dozhdikov, A.V. (2024). Enhancing State Policy Effectiveness in Cinema Through Machine Learning. *Nauka Televideniya—The Art and Science of Television*, 20 (2), 55–84. <https://doi.org/10.30628/1994-9529-2024-20.2-55-84>, <https://elibrary.ru/LXOADM>

# ENHANCING STATE POLICY EFFECTIVENESS IN CINEMA THROUGH MACHINE LEARNING

**Abstract.** A study was conducted on the distribution data of a range of Russian national films from 2004 to September 2023 using machine learning methods, with successful and unsuccessful films and patriotic projects considered separately. The study utilized the ensemble machine learning model HistGradientBoostingClassifier and a sequential fully connected three-layer neural network based on the TensorFlow library, along with basic methods of natural language processing. It was found that patriotic films exhibit distribution characteristics lower than the market average and significantly lag behind successful Russian cinema projects at the box office. The study demonstrated the possibility of accurately predicting film box office receipts, distribution characteristics, as well as selecting project parameters and the composition of its creative team to enhance distribution results and increase the reach of target audiences. The study also illustrated the use of language models, specifically through film annotations, to create effective film content. A conclusion was drawn regarding the need for a set of measures to implement state policy in the field of education, information technology, culture, and art, focusing on the creation of unified media franchises and umbrella brands encompassing films, TV series, computer and online games, music, and other products. The implementation of the research results will enhance the effectiveness of public policy and the return on public and private-public investments in the creative industries sector, positioning cinema and the creative industry as self-sustaining and income-generating sectors of the economy. Leveraging rental data from other markets will facilitate the creation of content with higher demand globally, serving as a soft power and cultural influence tool. The results of the study can aid in project selection and film project development by the Russian Cinema Fund, the Ministry of Culture of the Russian Federation, private Russian investors and film studios,

as well as authorities and organizations responsible for distributing Russian film content abroad.

**Keywords:** state policy, national cinema, natural language processing, language model, machine learning, artificial intelligence, neural network, soft power, cultural policy, Cinema Foundation of Russia, Ministry of Culture of the Russian Federation, Rossotrudnichestvo, Roscongress Foundation

## ВВЕДЕНИЕ

Для работы с длительными периодами времени и выработки стратегических решений, касающихся будущего политических систем, необходим важный компонент — политическая идеология с ее содержательными компонентами, транслируемыми в рамках государственной информационной политики как внутри страны, так и за ее пределами в контексте приложения «мягкой силы» (Павлова, 2021, с. 93) и осуществления «государственной культурной политики» (Эмих, 2015, с. 112), проведения конструктивной экспансии политической системы. Расширение применения искусственного интеллекта в данной отрасли позволит осуществить переход от аналитической к прогностической, творческой (генеративной) и проективной функциям. С помощью искусственного интеллекта возможно создание как отдельных художественных произведений, учебных материалов, так и франшиз, «зонтичных брендов», показывающих высокую эффективность и результативность по воздействию на целевые аудитории.

Основная **гипотеза** исследования связана с тем, что востребованность контента (кинопроизведения) и эффективность его воздействия на целевые аудитории обуславливается как содержанием, так и внешними признаками проекта (жанром, длительностью, возрастным рейтингом и другими), а также характеристиками творческой группы, задействованной в его создании. Следовательно, на основе данных проката прошлых периодов можно спрогнозировать успех или неуспех любого проекта заранее, до начала производственного процесса и возникновения основных затрат. Также возможен прогноз «успешности» кинопроекта у зрителей, исходя из его формального описания — развернутой аннотации, синопсиса и сценария, а впоследствии — аудиовизуального решения.

Доказательство выдвинутой гипотезы означает возможность создания высококачественной и востребованной целевыми аудиториями продукции,

нагруженной идеологическим содержанием, в данном случае — художественных фильмов, а впоследствии и литературных произведений, видеоигр, музыкальных клипов и других произведений сферы креативной индустрии за счет подбора оптимального сочетания характеристик творческой группы, выделения «критериев успеха» по результатам анализа исторических данных, сопоставления по содержанию с наиболее успешными в отрасли образцами. Также появится возможность для роста доли коммерчески прибыльных в стране и за ее пределами фильмов, кинематограф станет приносящей доход государству и бизнесу креативной отраслью, объектом для прямого и портфельного инвестирования.

## ОБЗОР ЛИТЕРАТУРЫ

Кинематограф — это высокоуровневый язык смыслов и образов, «язык программирования» для социальных систем. Это сопоставление основано на том, что досуговые предпочтения населения тесно связаны с ценностными и нравственными установками, формированием государственно-гражданской идентичности (Горшков, Шереги, 2019).

В мировом коммерческом кинематографе работа с автоматизированным «подбором» содержания кинопродукта и условиями его создания, прогнозированием сборов и проката поставлена на поток. Сложные исследовательские языковые модели и нейросети могут использоваться для выявления причин успешности сценариев и синопсисов — такие программные алгоритмы уже существуют в США, Европе (Murschetz et al, 2020) и особенно в Китае (Li et al, 2022), а также в других странах от Индии (Meenakshi et al, 2018; Chakraborty et al, 2019) и Шри-Ланки (Sivakumar et al, 2021) до Нигерии (Adecola et al, 2021) и Турции (Gürbüz et al, 2022).

Китайский кинорынок лидирует по числу исследований: для создания системы показателей в отношении кассовых сборов фильмов применяются экспертный опрос и метод Дельфи. При прогнозировании используются данные, характеризующие актеров, режиссера, сценариста, жанр проекта, награды творческой группы, ее известность, репутацию продюсера, график выхода и условия дистрибуции (Lu, 2019, p. 177). Но постепенно пальму первенства забирает машинное обучение (Abidi et al, 2020, p. 2) и глубокое обучение (Zhou et al, 2019, p. 1855).

Прогнозирование кинопроката осуществляется на основе работы с рецензиями и обзорами фильмов (Yoo & Kim, 2023, p. 95), микроблогами и неформальной коммуникацией (Zhao et al, 2022, p. 141). Прогностическая функция также реализуется с использованием изображений, глубоких мультимодальных визуальных функций, извлеченных из афиш фильмов и метаданных фильма (Madongo & Zhongjun, 2023, p. 1). Для оценки факторов, влияющих на кассовые сборы фильмов и поведение зрителей (Feng & Liu, 2020, p. 9), задействуются многофакторные и мультимодальные ансамблевые модели (Ni et al, 2022, p. 199). В последнее время отмечается усложнение методов в сторону применения именно ансамблевых моделей и объединения нескольких нейросетей. В частности, рядом исследователей используется глубокая нейронная сеть, которая объединяет признаки, извлеченные из постеров фильмов с помощью сверточной нейронной сети, а затем соответствующим образом разрабатывается набор новых генетических алгоритмов для анализа (Zhou & Yen, 2018, p. 1). Прогнозирование прокатных сборов фильмов, таким образом, снижает инвестиционный риск, поэтому оно имеет большое значение для киноинвесторов и социальной экономики (Chen & Dai, 2022).

Если в целом мировой пул публикаций и исследований в сфере машинного обучения применительно к кинематографу исчерпывается несколькими сотнями, то в современной Российской Федерации — считанными единицами, несмотря на активное внедрение методов data science в сфере финансов, ритейла, банковских услуг, логистики и доставки. Используются в основном традиционные социологические (опросы, анкетирование, фокус-группы) методы (Ноакк и др., 2015, с. 28) и оценка постфактум (пилотные группы зрителей после просмотра тизера или трейлера) (Ноакк и др., 2012, с. 17). Основной акцент сделан на математической обработке результатов оценки зрительских эмоций (Татарников, 2016). Применение методов машинного обучения и нейросетей касалось преимущественно иностранных кинорынков (Ясницкий и др., 2017, с. 449).

Вместе с тем с помощью машинного обучения мы можем выделить как формальные факторы успеха, определяющие удачу или неудачу проекта с высокой точностью (Дождииков, 2023), так и создать сложные ансамблевые модели-классификаторы, позволяющие определить не только класс успеха/неуспеха проекта, но и более сложную классификацию, применив алгоритмы регрессии для определения зрительского рейтинга, количества просмотров, величины сборов и окупаемости (Дождииков, 2024).

## МАТЕРИАЛЫ И МЕТОДЫ

Объектом исследования выступают прокатные результаты и доступные исследователю открытые данные российских национальных фильмов с февраля 2004 по сентябрь 2023 года — это 1683 проекта, выпущенных в прокат, без учета повторов в 2022 году (на фоне ухода иностранных кинодистрибьютеров с российского кинорынка), а также альманахов, отдельных документальных и короткометражных картин, как и картин, не вышедших в открытый прокат или вышедших сразу на телевидении, Интернет-платформах, стриминговых сервисах и в онлайн-кинотеатрах<sup>1</sup>.

В исследовании использовались открытые официальные и неофициальные базы данных современного российского проката<sup>2</sup>, предоставляющие контент без ограничений на его использование в научных, образовательных и культурных целях,<sup>3</sup> с последующей ручной и автоматизированной проверкой и сопоставлением данных из разных источников, сведением в единый датасет.

С февраля 2004 года по сентябрь 2023 года 1683 кинокартины заработали 160,7 миллиардов рублей при 727,2 миллионов просмотров и совокупных открытых расходах на производство 163,3 миллиарда. Поскольку примерно по 1/3 всех кинопроектов данные о бюджете отсутствуют, для расчетов использовались оценочные значения на основе медианных показателей. Неполнота данных создает проблемы для получения точных прогнозов по сборам и числу зрителей с использованием алгоритмов регрессии.

За годовой период с 1 января по 31 декабря 2023 года в официальный российский кинопрокат вышло 192 проекта<sup>4</sup>, включая фильмы-копродукцию. Совокупные сборы кинотеатров превысили 40 млрд рублей<sup>5</sup>, из которых

<sup>1</sup> Данные за 2023-2024 годы будут использованы для валидации разработанных моделей в продолжении исследования.

<sup>2</sup> Автор выражает признательность владельцам и администраторам открытых новостных ресурсов и порталов данных: Kinopoisk.ru, kinometro.ru, Afisha.ru, Film.ru, kinobusiness.com, kino-teatr.ru и kinopoisk\_dev.

<sup>3</sup> ГК РФ Статья 1274. Свободное использование произведения в информационных, научных, учебных или культурных целях «Гражданский кодекс Российской Федерации (часть четвертая)» от 18.12.2006 N 230-ФЗ (ред. от 13.06.2023, с изм. от 14.12.2023) (с изм. и доп., вступ. в силу с 29.06.2023).

<sup>4</sup> Лавров, С. (2024). Российское кино — итоги 2023 года. *Синемалекс*. 08.01.2024 (Электронный ресурс). URL: <https://cinemaplex.ru/2024/01/08/rossijskoe-kino-itogi-2023-goda-kinoproizvodstvo-proczvetaet-oficialnyj-kinopokaz-chahnet.html> (дата обращения: 25.03.2024).

<sup>5</sup> Пресс-конференция, посвященная российскому кинематографу в 2023 году. (2023). *Фонд кино*. 28.12.2023 (Электронный ресурс). URL: <https://www.fond-kino.ru/news/press-konferencia-posvasenna-rossijskomu-kinematografu-v-2023-godu-28-12-2023/> (дата обращения: 25.03.2024).

примерно 70%<sup>6</sup> пришлось на национальные фильмы и проекты совместного производства. Но в целом для государства и частных инвесторов российский кинематограф продолжает быть убыточным. У данной ситуации есть объективная причина: для окупаемости необходима страновая аудитория не менее 500 миллионов человек.

По степени распространенности на планете, по количеству владеющих языком людей российский кинорынок в 2022–2023 годах по разным источникам<sup>7</sup> занимал восьмое-девятое место в мире — это 255 млн человек, то есть в 2 раза меньше, чем необходимо для создания самостоятельного самоокупающегося кинорынка. По этой же причине условные «блокбастеры» с бюджетом свыше 750 миллионов рублей, как правило, оказываются не рентабельны. Единственное исключение за исследуемый период времени — фильм «Чебурашка»<sup>8</sup>. Усредненные данные проката приведены в Таблице 1.

Таблица 1\*

Усредненные данные российского проката 1683 кинокартин

Table 1\*

Average Data for Russian Distribution of 1683 Films

Параметр / Parameter	Российский кинематограф в целом / Russian films in general	Патриотический кинематограф / Patriotic films	Успешные в прокате проекты / Successful projects in terms of the box office	Размеренность / Regularity
Количество экранов в прокате / Number of screens	521	774,4	1089,8	экраны / screens
Бюджет / Budget	141	308	117	млн руб. / mln. Rub.
Длительность картины / Film duration	96	107	97,7	минуты / minutes

<sup>6</sup> Корнацкий, Н. (2023). Доля сборов российского кино по итогам 2023 года превысила 70%. *Ведомости*. 28.12.2023 (Электронный ресурс). URL: [https://www.vedomosti.ru/media/articles/2023/12/28/1013456-dolya-sborov-rossiiskogo-kino?from=popular\\_search\\_1](https://www.vedomosti.ru/media/articles/2023/12/28/1013456-dolya-sborov-rossiiskogo-kino?from=popular_search_1) (дата обращения: 25.03.2024).

<sup>7</sup> Как изменилось положение русского языка в мире за последние 30 лет (2022). *РБК.Тренды*. 31.03.2022. (Электронный ресурс). URL: <https://trends.rbc.ru/trends/social/624591cc9a7947d35bf12bfc> (дата обращения: 25.03.2024).

<sup>8</sup> Дождиков, А.В. (2023). Успех «Чебурашки» можно повторить с помощью ИИ. И не только в киноиндустрии!. *PLUSworld*. 22.08.2023 (Электронный ресурс). URL: <https://plusworld.ru/journal/2023/plus-3-2023/uspekhe-cheburashki-mozhno-povtorit-s-pomoshchyu-ii-i-ne-tolko-kinoindustrii/> (дата обращения: 25.03.2024).



Сборы на кинокартину / <i>Box office revenue per film</i>	94,5	220	488	млн руб. / <i>mln. Rub.</i>
Количество просмотров фильма / <i>Number of film views</i>	427,6	1206	2199	тыс. просмотров / <i>thous. views</i>
Сборы (наработка) на экран / <i>Box office revenue per screen</i>	113,5	226	435	тыс. руб. / <i>thous. Rub.</i>
Просмотры на экран / <i>Views per screen</i>	613	1300	2269	просмотры / <i>views</i>
Соотношение сборы / бюджет / <i>Box office-to-budget ratio</i>	1,084	0,825	4,77	
Рейтинг Кинопоиска / <i>Kinopoisk rating</i>	5,79	5,86	5,78	

\* *Источник данных: составлено автором на основе первичных открытых данных информационных систем / Data source: compiled by the author from primary open data available at various information systems.*

Только 180 кинокартин из 1683 можно назвать окупившимися в прокате (10,7%), т. е. собравшими два своих бюджета и более. Всего 161 картину (9,5%) можно отнести к условному «патриотическому» кинематографу, с учетом их жанровой (как правило, историческая, военная драма) и содержательной направленности.

Уже на данном предварительном этапе количественного анализа очевидно, что «патриотический кинематограф» существенно уступал остальным проектам. Условный дорогостоящий «патриотический блокбастер» имеет большую длительность (почти на 10 минут) и бюджет почти в три раза выше самой коммерчески окупаемой продукции. По соотношению сборы/бюджет он проигрывает типовому российскому фильму (0,825 против 1,084). В то время как средний бюджет успешных в прокате картин — всего 117 миллионов, среднее соотношение сборы/бюджет 4,77 при фактически одинаковом зрительском рейтинге. Отметим сразу, что коэффициент корреляции зрительского рейтинга, сборов и просмотров, а также количества наград и призов кинофестивалей у режиссеров и продюсеров всего 0,053, то есть какая-либо значимая связь между «наградами», «рейтингами», «экспертными оценками» и финансовыми результатами в прокате отсутствует. Соответственно, инвестиционные решения или целевые государственные заказы на основе субъективного экспертного мнения в ряде случаев убыточны и неэффективны.

Таким образом, «патриотический кинематограф» долгое время был финансово неэффективным, притом что количество просмотров и другие прокатные характеристики, благодаря активным информационным кампаниям и продвижению в СМИ, превышают среднероссийский уровень.

Существенное отличие данного поджанра кинематографа от остального — это доминирование темы войны (как правило, Великой Отечественной) и специфический жанр — военная или историческая драма, изредка — боевик или «экшен». В то время как успешное в российском прокате кино — это, как правило, сказка, анимация, комедия, мелодрама, музыкальный фильм. (Подробную информацию по показателям и жанрам, выбор которых влияет на успех/неуспех проекта, см.: Дождииков, 2024.)

Общая характеристика российского кинематографа — его «локальность», он «свой» для современного зрителя, но только в пределах Российской Федерации и стран постсоветского пространства. Национальный кинематограф отражает историю страны и ее специфическое культурное наследие. Как следствие, отсутствуют значимые научно-фантастические проекты или фильмы жанра «экшен», востребованные целевыми аудиториями на Западе и Востоке, что сокращает возможности для осуществления культурно-гуманитарного воздействия и проецирования «мягкой силы».

Одним из предполагаемых направлений развития таких возможностей является смена формата, некоторых прокатных характеристик и жанра. К примеру, комедия вместо обычной драмы позволяет не только продвигать «традиционные» ценности, но и с гораздо большим успехом дискредитировать «нетрадиционные»: такие поджанры, как интеллектуальная «черная» комедия или комедия абсурда, дают возможность работы с иностранными западными аудиториями, а сказка, анимация, комикс, боевик — с восточными целевыми аудиториями.

Для того чтобы доказать возможность с определенной точностью прогнозировать результаты кинопроката по внешним показателям, включая прокатные характеристики и данные творческой группы, использованы параметры классификации «точность» (англ. — accuracy, иногда возможен перевод как «правильность») и количественная интерпретация ROC (англ. — Receiver Operating Characteristic<sup>9</sup>, рабочая характеристика приемника), которая дает показатель AUC (англ. — Area Under Curve, площадь под кривой) — площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций. Чем выше показатель AUC, тем качественнее классификатор, при этом значение 0,5 соответствует случайному гаданию.

<sup>9</sup> Receiver Operating Characteristic (Электронный ресурс). URL: [https://link.springer.com/referenceworkentry/10.1007/978-0-387-39940-9\\_569](https://link.springer.com/referenceworkentry/10.1007/978-0-387-39940-9_569) (дата обращения: 25.03.2024).

Значение менее 0,5 показывает, что классификатор действует с точностью до наоборот.

Для выяснения возможности прогнозирования результатов кинопроката по внешним характеристикам используем 26-факторную модель, включающую прокатные данные, исторические данные по жанру, режиссеру и сценаристам проекта:

- week — порядковый номер недели выхода фильма с начала года;
- month — порядковый номер месяца выхода фильма с начала года;
- screens — начальное количество экранов проката (экранных копий);
- budget — бюджет фильма (в миллионах рублей);
- age\_R — возрастной рейтинг фильма от 0+ до 18+;
- time — общая длительность фильма в минутах;
- genre\_box\_buget — среднее соотношение сборы/бюджет по жанру;
- genre\_avr\_kinopoisk\_R — средний зрительский рейтинг Кинопоиска по жанру;
- genre\_avr\_box — средние сборы по жанру;
- genre\_avr\_views — среднее количество просмотров для жанра проекта;
- genre2\_box\_buget — сборы/ бюджет для поджанра проекта;
- genre2\_avr\_kinopoisk\_R — средний рейтинг для поджанра проекта;
- genre2\_avr\_box — средние сборы для поджанра проекта;
- genre2\_avr\_views — средние просмотры для поджанра проекта;
- dir\_box\_buget — среднее соотношение сборы/бюджет для режиссера;
- dir\_avr\_kinopoisk\_R — средний рейтинг режиссера на Кинопоиске;
- dir\_avr\_box — средняя величина сборов проектов данного режиссера;
- dir\_avr\_views — средняя величина просмотров фильмов режиссера;
- skr1\_box\_buget — среднее соотношение сборы/бюджет для сценариста;
- skr1\_avr\_kinopoisk\_R — средний зрительский рейтинг для сценариста;
- skr1\_avr\_box — средняя величина сборов по проектам сценариста;
- skr1\_avr\_views — среднее количество просмотров проектов сценариста;
- skr2\_box\_buget — соотношение сборы/бюджет второго сценариста;
- skr2\_avr\_kinopoisk\_R — средний зрительский рейтинг второго сценариста;

skr2\_avr\_box — среднее количество сборов прехтов второго сценариста;

skr2\_avr\_views — среднее количество просмотров второго сценариста.

Отметим, что пункты 4–26 известны задолго до начала съемок проекта — на этапе продюсерского питчинга или представления проекта кинокартины частному инвестору либо государственной экспертной комиссии, а пункты 1–4 известны до даты его выхода в прокат.

На основании этих данных прогнозируется класс проекта: 0 — провалился в прокате, 1 — не окупился, 2 — окупился, 3 — принес прибыль свыше 100% производственного бюджета. Построим матрицу корреляций для указанных значений для обнаружения возможных связей между рассматриваемыми признаками (см. рис. 1).

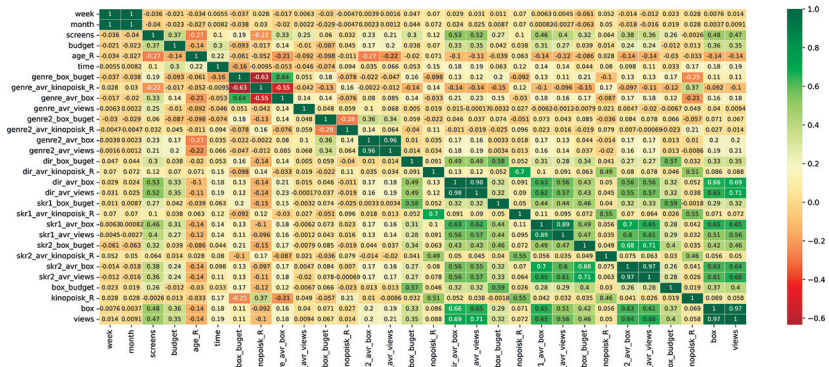


Рис. 1 Матрица корреляций оценочных признаков

Fig. 1 Evaluation feature correlation matrix\*\*

\*\* Здесь и далее источник являются данные, полученные автором по результатам реализации разработанного программного кода / Hereinafter, the data source is derived from the author's analysis of the outcomes resulting from the execution of the developed program code<sup>10</sup>

В качестве отдельной аномалии отметим достаточно заметную отрицательную корреляционную связь между genre\_avr\_kinopoisk\_R и genre\_box\_budget, genre\_avr\_box, что может говорить о том, что высокий

<sup>10</sup> Датасет по кинофильмам и исходный программный код для 26-факторной модели, применяющийся в обработке данных для количественных расчетов (классификация и регрессия) представлен в открытом обновляемом репозитории на github.com (Электронный ресурс). URL: <https://github.com/AntonDzhidkov/movie-box-office-prediction> (дата обращения: 11.06.2024).

зрительский рейтинг как раз не предопределяет окупаемость фильма в прокате (и наоборот) — это может быть связано с несовершенством методики подсчета зрительского рейтинга «Кинопоиска»: 1) авторское кино для узких целевых групп, не окупающееся в прокате, при меньшем суммарном количестве оценок получает заведомо большие баллы, демонстрируя сплоченность и солидарность оценщиков: выборка голосующих не репрезентативна всей генеральной совокупности целевой аудитории; 2) ряд кинокартин<sup>11</sup> на острополитические темы имел заведомо «протестный» пул голосующих, в том числе и из-за рубежа, также с высоким уровнем солидарности и неприятия — эта группа голосующих также не репрезентативна по отношению к генеральной совокупности зрительской аудитории в России; 3) остается актуальной проблема манипуляций зрительскими рейтингами как в случае с организованными в соцсетях кампаниями по дискредитации и накрутке рейтинга, так и в случае применения алгоритмов повышения популярности платного киноконтента, известных в сообществе киноаналитиков как минимум с 2015 года<sup>12</sup>. Поэтому показатели «окупаемость», «сборы» и «просмотры» представляются более объективными, чем «зрительские рейтинги».

Определенный фактор — это высокая взаимозависимость между признаками `week` и `month`, между `genre_avr_box` и `genre_avr_views`, `dir_avr_box` и `dir_avr_views`, `skr1_avr_box` и `skr1_avr_views`, `skr2_avr_box` и `skr2_avr_views` и, соответственно, `box` и `views`. При дальнейшей работе эти признаки можно исключать или использовать алгоритмы машинного обучения, позволяющие работать с такими данными. Поэтому одним из наиболее подходящих выбран `HistGradientBoostingClassifier`; возможно применение такого инструмента, как `CatBoost` от «Яндекса».

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Используем стандартную ансамблевую модель машинного обучения `HistGradientBoostingClassifier` и нейросеть, построенную на архитектуре

---

<sup>11</sup> Характерный пример фильм «Свидетель» Давида Дадунашвили. URL: <https://www.kinopoisk.ru/film/5332755/> (дата обращения: 11.06.2024).

<sup>12</sup> Новый алгоритм расчета рейтинга «Кинопоиска» отдает предпочтение фильмам с платным просмотром. Мини-расследование. URL: <https://habr.com/ru/articles/368759/> (дата обращения: 11.06.2024).

библиотеки TensorFlow<sup>13</sup> — последовательную полносвязную сеть из трех слоев: 1) полносвязный слой (Dense) с функцией активации ReLU<sup>14</sup>, принимающий «на входе» тензор размерности 26 (по количеству признаков); 2) полносвязный слой (Dense) с 64 нейронами и функцией активации ReLU; 3) полносвязный слой (Dense) с 4 нейронами и функцией активации Softmax<sup>15</sup> «на выходе»<sup>16</sup>.

В первом случае работаем с несимметричными данными классов, разделив выборку на трениговую и тестовую в пропорции 7 к 3. Во втором случае постараемся повысить точность и другие показатели модели, используя увеличение выборки с использованием библиотеки imbalanced-learn<sup>17</sup> и метода RandomOverSampler<sup>18</sup> — с помощью синтетических данных сделаем количество экземпляров каждого из четырех классов примерно равными. Использование синтетических данных оправдано для обучения моделей в случае дисбаланса классов и ряда других недостатков исходного датасета.

Для четырехклассовой классификации наилучшие результаты по параметрам accuracy=0,8878 и ROC\_AUC\_score=0,9611 показала модель HistGradientBoostingClassifier. С подобранными методом GridSearchCV<sup>19</sup> гиперпараметрами (learning\_rate=0.1, max\_depth=4, max\_iter=80) получена точность=0,8918 и ROC\_AUCscore=0,9653 (см. Таблицу 2).

---

<sup>13</sup> An end-to-end platform for machine learning (Электронный ресурс). URL: <https://www.tensorflow.org/> (дата обращения: 25.03.2024).

<sup>14</sup> A Gentle Introduction to the Rectified Linear Unit (ReLU) by Jason Brownlee on August 20, 2020 in Deep Learning Performance (Электронный ресурс). URL: <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/> (дата обращения: 25.03.2024).

<sup>15</sup> Выбор слоя активации в нейронных сетях: как правильно выбрать для вашей задачи (Электронный ресурс). URL: <https://habr.com/ru/articles/727506/> (дата обращения: 25.03.2024).

<sup>16</sup> Автором были использованы разные комбинации. В репозитории проекта представлены также вариант 3-слойного перцептрона для четырехклассовой классификации и 8-слойного перцептрона в варианте для регрессии.

<sup>17</sup> Imbalanced-learn documentation Jan 24, 2024 Version: 0.12.0 (Электронный ресурс). URL: <https://imbalanced-learn.org/stable/> (дата обращения: 25.03.2024).

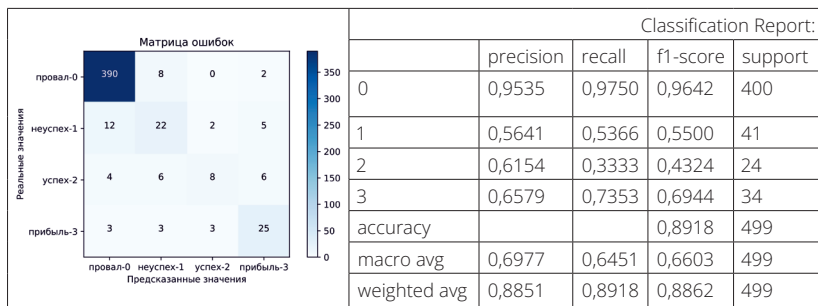
<sup>18</sup> RandomOverSampler (Электронный ресурс). URL: [https://imbalanced-learn.org/stable/references/generated/imblearn.over\\_sampling.RandomOverSampler.html](https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.RandomOverSampler.html) (дата обращения: 25.03.2024).

<sup>19</sup> Sklearn.model\_selection.GridSearchCV (Электронный ресурс). URL: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html) (дата обращения: 25.03.2024).

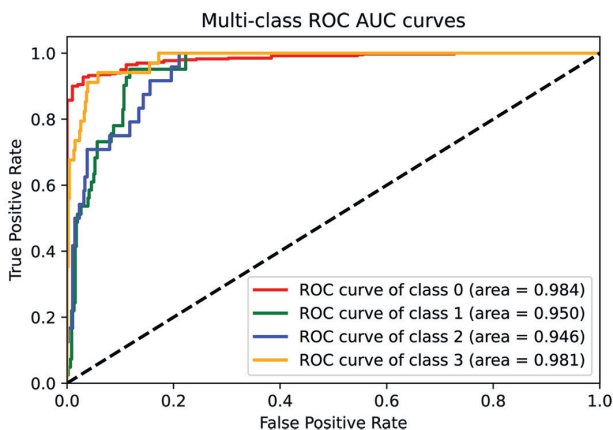
Таблица 2

Результат классификации при помощи HistGradientBoostingClassifier

Table 2

*HistGradientBoostingClassifier Classification Results*

При переводе меток классов в формат *one-vs-all*<sup>20</sup> получены *ROC\_AUC\_score* для каждого класса (рис. 2). По данному критерию показатели отбора намного превышают данные «случайного» отбора и на исторических данных превышают показатели отбора любой конкурсной комиссии по критерию «успех в прокате».

Рис. 2. *ROC\_AUC\_score* 4-классовой классификации модели машинного обученияFig. 2. *ROC\_AUC\_score* for the 4-class machine learning model classification

<sup>20</sup> Sklearn.multiclass.OneVsRestClassifier (Электронный ресурс). URL: <https://scikit-learn.org/stable/modules/generated/sklearn.multiclass.OneVsRestClassifier.html> (дата обращения: 25.03.2024).

Возможность прогнозирования кинопроката доказана, однако следующие метрики недостаточно убедительны:

Accuracy (точность) — количество правильно предсказанных результатов классификации;

Precision (правильность) — количество предсказанных положительных результатов (истинно положительных), деленное на все предсказанные положительные результаты (истинно положительный + ложноположительный результат);

Recall (отклик, чувствительность) — количество предсказанных положительных результатов (истинно положительных), деленное на общее количество положительных результатов (истинно положительный + ложноотрицательный);

f1-score (оценка производительности модели) — среднее гармоническое значение точности и отклика.

Предложенная модель машинного обучения недостаточно качественно отделяет потенциально кассовые проекты от просто успешных, провалившиеся в прокате от не окупившихся. Для повышения точности прогнозов проведем балансировку предсказываемых классов с помощью синтетических данных и используем отмеченную ранее последовательную полносвязную трехслойную нейросеть, скомпилируем ее при помощи следующих параметров: optimizer="adam"<sup>21</sup>, loss="categorical\_crossentropy"<sup>22</sup>, metrics=("accuracy")<sup>23</sup>.

Мы обучим данную нейросеть в течение 300 эпох. Чтобы не допустить ее переобучения и ухудшения характеристик, используем ModelCheckpoint<sup>24</sup>, сохраняющую лучшие параметры настройки нейросети. Получим историю обучения и выберем эпоху обучения, где производительность нейросети оптимальна. На графике функции потерь (loss) (рис. 3), очевидно, что функция потерь после 200 эпохи стала минимальной в отношении данных как на тренинговой, так и на валидационной выборках. Дальнейшее обучение и тренировка нейросети становится нецелесообразным.

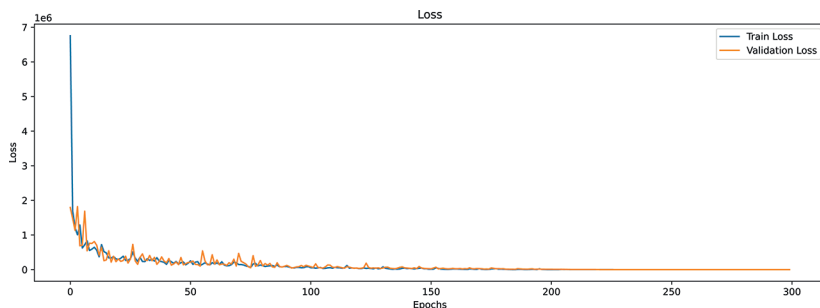
<sup>21</sup> Adaptive Moment Estimation. Tf.keras.optimizers.legacy.Adam (Электронный ресурс). URL: [https://www.tensorflow.org/api\\_docs/python/tf/keras/optimizers/legacy/Adam](https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/legacy/Adam) (дата обращения: 25.03.2024).

<sup>22</sup> Tf.keras.metrics.categorical\_crossentropy (Электронный ресурс). URL: [https://www.tensorflow.org/api\\_docs/python/tf/keras/metrics/categorical\\_crossentropy](https://www.tensorflow.org/api_docs/python/tf/keras/metrics/categorical_crossentropy) (дата обращения: 25.03.2024).

<sup>23</sup> Accuracy metrics (Электронный ресурс). URL: [https://keras.io/api/metrics/accuracy\\_metrics/](https://keras.io/api/metrics/accuracy_metrics/)

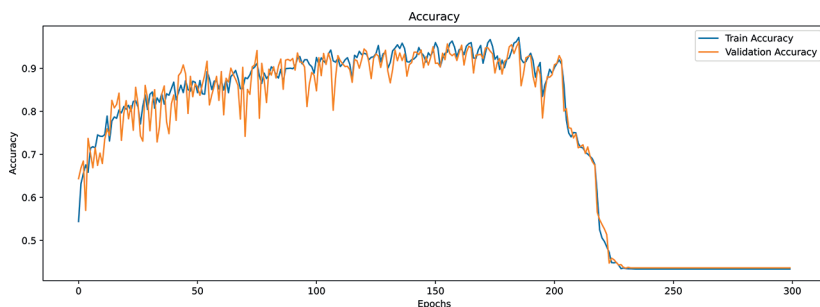
<sup>24</sup> Tf.keras.callbacks.ModelCheckpoint (Электронный ресурс). URL: [https://www.tensorflow.org/api\\_docs/python/tf/keras/callbacks/ModelCheckpoint](https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/ModelCheckpoint) (дата обращения: 25.03.2024).





**Рис. 3. График обучения нейросети: значение функции потерь от эпохи обучения**  
*Fig. 3. Neural network training graph: the loss function value of the training epoch*

График точности (accuracy) (рис. 4) показывает, что переобучение началось после 170 эпохи, что для исследования не является принципиальной проблемой, т. к. мы уже зафиксировали нейросеть в ее состоянии с максимальными показателями производительности и качества и будем использовать их и далее.



**Рис. 4. График обучения нейросети: значение «точности» на тренинговой и тестовой выборках**  
*Fig. 4. Neural network training graph: the “accuracy” value on the training and testing sets*

На тестовом наборе данных нейросеть показывает высокие результаты в распознавании будущего успеха кинокартин по показателям accuracy, precision, recall, f1-score. Значения ROC\_AUC также близки к максимально возможным.

Однако эти показатели достигнуты с использованием синтетических данных на сбалансированных классах: к таким «высоким» результатам следует относиться с осторожностью.

Результат классификации при помощи нейросети

Table 3

Classification result using a neural network

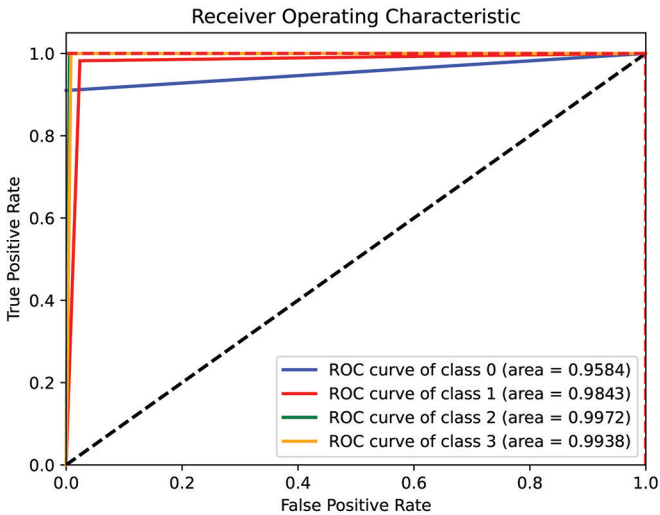
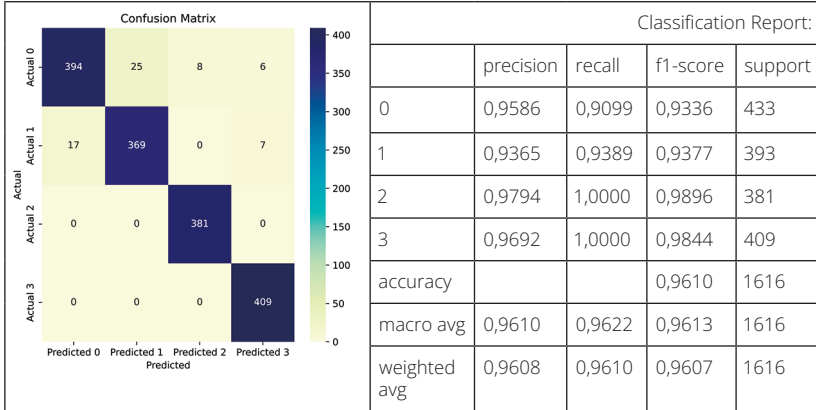


Рис. 5. ROC\_AUC\_score 4-классовой классификации трехслойной нейросети прямого распространения

Fig. 5. ROC\_AUC\_score of the 4-class classification of a three-layer feedforward neural network

Точность прогнозов ансамблевой модели машинного обучения на несбалансированной по классам тестовой выборке составила 0,892, нейросети — 0,961 (на сбалансированной по классам с помощью синтетических данных тестовой выборке). В первом случае ROC\_AUC составил 0,965, во втором — 0,985. Продолжение исследования предполагает использование данных новейшего периода (2023–2024) для оценки эффективности созданных моделей.

Другое направление исследования может быть связано с семантической оценкой аннотаций, синопсисов и сценариев российских национальных кинопроектов с использованием нейросетей и больших языковых моделей. Отметим наличие различий в содержании между категориями «успешное в прокате кино» и «патриотический кинематограф». Для этого используем языковую модель spaCy<sup>25</sup> для русского языка (“ru\_core\_news\_sm”), проведем токенизацию текста расширенных аннотаций 180 успешных кинопроектов и 161 «патриотических» с помощью NLTK<sup>26</sup>, очистку, лемматизацию и фильтрацию токенов, удаление нерелевантных символов из слова, пропуск слов с длиной меньше 3 символов. С использованием CountVectorizer<sup>27</sup> произведем векторизацию текста и преобразование в базовую модель вида «мешок слов». Для визуализации «мешка» на основе словаря с частотами используем экземпляр класса WordCloud<sup>28</sup>. Результаты анализа частоты встречаемости слов для двух данных категорий представлены на рис. 6 и рис. 7.

---

<sup>25</sup> Обработка и анализ естественного языка с помощью Python-библиотеки spaCy. 21 авг 2023 (Электронный ресурс). URL: <https://habr.com/ru/companies/otus/articles/755584/> (дата обращения: 25.03.2024).

<sup>26</sup> Natural Language Toolkit (Электронный ресурс). URL: <https://www.nltk.org/> (дата обращения: 25.03.2024).

<sup>27</sup> Sklearn.feature\_extraction.text.CountVectorizer (Электронный ресурс). URL: [https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_extraction.text.CountVectorizer.html](https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html) (дата обращения: 25.03.2024).

<sup>28</sup> WordCloud for Python documentation (Электронный ресурс). URL: [https://amueller.github.io/word\\_cloud/](https://amueller.github.io/word_cloud/) (дата обращения: 25.03.2024).



Рис. 6. Частотный анализ описаний успешных кинокартин

Fig. 6. Frequency analysis of successful film descriptions



Рис. 7. Частотный анализ описаний патриотических кинокартин

Fig. 7. Frequency analysis of patriotic film descriptions

Для определения семантической близости между двумя наборами слов воспользуемся такими мерами расстояния, как косинусное расстояние (Потапенко, 2019, с. 107; Каряева и др., с. 728), расстояние Жаккара (Посов, Допира, 2019). Они позволяют оценить, насколько много общего содержится в наборах описаний кинокартин.

Для расчета косинусного расстояния используется *CountVectorizer* из библиотеки *scikit-learn* и *cosine\_distances*<sup>29</sup> из подраздела *sklearn.metrics*.

<sup>29</sup> Scikit-learn. (n.d.). Cosine\_distances. URL: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.cosine\\_distances.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.cosine_distances.html) (дата обращения: 25.03.2024).

*pairwise*, а для расчета расстояния Жаккара используется функция *jaccard\_similarity*<sup>30</sup>, которая принимает на вход два множества слов.

Применительно к двум корпусам аннотаций косинусное расстояние в размере 0,21876 говорит о том, что тексты имеют некоторое сходство в терминах слов и в их взаимном расположении в пространстве, но не являются похожими. Угол между векторами, представляющими эти тексты, составляет примерно 77 градусов, что указывает на относительную разницу между ними. Расстояние Жаккара, равное 0,15385, означает, что между множествами слов в обоих текстах есть некоторое пересечение, но оно не является очень значительным. Это расстояние измеряет сходство между множествами элементов, поэтому оно указывает на небольшое количество общих слов между текстами. При большем сходстве в описаниях — больше шансов на успех картины. Данное направление исследования также является перспективным в контексте сравнения двух корпусов, формализованных данных и других результатов, получаемых с помощью моделей *ru\_core\_news\_lg*, *ru\_core\_news\_md*, *ru\_core\_news\_sm*<sup>31</sup> и других, и будет продолжено в последующих публикациях.

## ВЫВОДЫ

В ходе исследования доказана возможность подбора содержательных элементов описания кинопроекта для улучшения характеристик кинопроката: чем меньше расхождение между «эталонным» набором описаний успешных фильмов и новым описанием, тем больше шансов на успех у нового кинофильма. Данный подход отнюдь не ограничивает просторы для творческого замысла автора, который должен всего лишь учитывать действующие тенденции, тренды и интересы целевой аудитории, задействуя в процессе восприятия глубинные психологические архетипы зрительской аудитории, которые достаточно консервативны и меняются очень медленно.

На основе построения более сложной модели исследования с использованием средств обработки натурального языка появится возможность анализа не только аннотаций, но и сценариев кинофильмов для

<sup>30</sup> Scikit-learn. (n.d.). Jaccard\_score. URL: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.jaccard\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.jaccard_score.html) (дата обращения: 25.03.2024).

<sup>31</sup> Хабр. (2021, 2 июля). Сравнение распознавания сущностей русского языка в spaCy 2 и spaCy 3. URL: <https://habr.com/ru/sandbox/156312/> (дата обращения: 01.07.2024).

определения их потенциальной успешности/неуспешности с последующим подбором наиболее удачных элементов сюжета: современный сценарий в так называемом «американском» формате прекрасно поддается формализации и машинной обработке, а обучение предполагаемой языковой модели на корпусе успешных и неуспешных сценариев российских кинофильмов позволит создавать более качественную кинопродукцию, повышать производительность и качество труда и результатов участников кинопроизводственной деятельности.

Использование машинного обучения, глубокого обучения и больших языковых моделей также позволит реализовать принципы сторителлинга при создании «зонтичных медиафраншиз» с государственным и частно-государственным участием, органически вплетая в повествование вопросы государственно-патриотического воспитания, духовно-нравственного развития без использования методов и методик директивной подачи идей и смыслов, вызывающих лишь отторжение целевых аудиторий. Результатом будет насыщение киноконтента брендами национальной экономики — от туристической отрасли до продукции и услуг, оказываемых крупными государственными, частно-государственными компаниями и корпорациями. Появится возможность более эффективно продвигать национальные бренды и символы на международной арене и внутри страны.

Основная гипотеза исследования также доказана. Результаты кинопроката можно определить заранее, до момента выхода фильма на большие экраны, в части признаков — до начала съемок самого фильма, на этапе продюсерского питчинга перед инвесторами, в «Фонде Кино»<sup>32</sup> или Министерстве культуры Российской Федерации. Предварительные результаты исследования опубликованы<sup>33</sup> для сообщества специалистов в data science<sup>34</sup>, инициирована дискуссия по уточнению методов, поиску возможных ошибок (использование «синтетических данных», возможности тюнинга моделей) и совершенствованию применяемых методик оценки.

С помощью оценки характеристик будущего кинофильма, предшествовавшей истории и исторических данных творческой группы можно

---

<sup>32</sup> В соответствии с постановлением Правительства Российской Федерации от 27 августа 2018 года № 1001 «О Федеральном фонде социальной и экономической поддержки отечественной кинематографии» (в редакции постановления Правительства Российской Федерации от 22 октября 2020 года № 1718, постановления Правительства Российской Федерации от 13 июня 2024 г. N 797).

<sup>33</sup> Rain, A. (2024, 27 мая). Прогнозируем результаты российского кинопроката с помощью ML. URL: <https://habr.com/ru/articles/817471/> (дата обращения: 01.07.2024).

<sup>34</sup> Rain, A. (2024, 20 июня). Кино, финансы и data science. URL: <https://habr.com/ru/articles/823368/> (дата обращения: 01.07.2024).

с высокой точностью определять будущий успех/неуспех кинофильма и востребованность его у зрителей.

Следующий этап — это подбор создания эффективного контента с помощью машинного обучения. Таким образом, появится возможность создавать кинофильмы, отражающие в т. ч. принципиальную идеологическую позицию государства, которые будут востребованными у целевой аудитории вне зависимости от идеологического контекста (большие языковые модели впоследствии смогут помогать создавать произведения с учетом необходимых смыслов и установок).

Использование исторических данных проката других локальных кинорынков (например, стран БРИКС) позволит генерировать проекты, в большей степени привлекающие интерес целевых аудиторий иностранных государств.

В рамках реализации государственной политики в сфере культуры и искусства использование искусственного интеллекта будет способствовать более эффективному принятию решений, адаптируя систему их выработки к новым вызовам, особенно «в условиях санкций и других ограничений» (Добротворская, 2019, с. 297), и, одновременно, интенсифицируя развитие новых отраслей не только в федеральном центре, но и в регионах (Астафьев и др., 2021). В целом это окажет системное содействие реализации стратегических документов по развитию страны (Михайлова, 2023, с. 667), в т. ч. Указа Президента Российской Федерации от 07.05.2024 № 309 «О национальных целях развития Российской Федерации на период до 2030 года и на перспективу до 2036 года» в части целей «б», «д», «е».

Точное прогнозирование результатов кинопроката даст возможность задействовать различные варианты поддержки и финансирования кинематографа — более чем 120 видов и форм (Юсупова, 2021, с. 275) из мировой практики. Ежегодный вывод на рынок 20–30 качественных кинокартин с улучшенными прокатными характеристиками в рамках 7–10 глобальных кинофраншиз, включающих видеоигры, сериалы, музыкальные клипы, реалити-шоу и другие направления, их трансляция на зарубежные целевые аудитории станут эффективным инструментом государственной культурной политики и проекции «мягкой силы» на постсоветском пространстве и за его пределами.

## ЛИТЕРАТУРА

1. Астафьев, С.А., Хомкалов, Г.В., Толстоухова, И.С. (2021). Креативная экономика как ключевой элемент устойчивого развития территорий. *Baikal Research Journal*, 12 (3), <https://www.elibrary.ru/KMRAXB>, [https://doi.org/10.17150/2411-6262.2021.12\(3\).16](https://doi.org/10.17150/2411-6262.2021.12(3).16)
2. Горшков, М.К., Шереги, Ф.Э. (2019). *Российская молодежь в контексте социологического анализа*. М: Федеральный научно-исследовательский социологический центр Российской академии наук. 263 с. <https://www.elibrary.ru/ZRPVCL>, <https://doi.org/10.19181/monogr.978-5-89697-317-1>
3. Добротворская, Д.А. (2023). Роль государственной поддержки в развитии креативных индустрий в условиях санкций и ограничений. И.М. Степнов, Ю.А. Ковальчук (ред.) *Управление активами — 2023: Бизнес-модели в эпоху изменения делового климата*. Москва: Федеральное государственное бюджетное учреждение науки Институт проблем рынка Российской академии наук, 297–301. <https://www.elibrary.ru/LPNIHO>
4. Дождииков, А.В. (2023). Прогнозирование результатов кинопроката с помощью машинного обучения. *Вопросы теоретической экономики*, 4 (21), 93–114. <https://www.elibrary.ru/PNLSGN>, [https://doi.org/10.52342/2587-7666VTE\\_2023\\_4\\_93\\_114](https://doi.org/10.52342/2587-7666VTE_2023_4_93_114)
5. Дождииков, А.В. (2024). Определение инвестиционного успеха и его факторов для российского кино в прокате с помощью машинного обучения. *Финансы: теория и практика*, 28 (1), 188–203. <https://www.elibrary.ru/PSGUET>, <https://doi.org/10.26794/25875671-2024-28-1-188-203>, <https://financetp.fa.ru/jour/article/view/2690>
6. Каряева, М.С., Браславский, П.И., Соколов, В.А. (2018). Векторное представление слов с семантическими отношениями: экспериментальные наблюдения. *Моделирование и анализ информационных систем*, 25.6 (78), 726–733. <https://www.elibrary.ru/YQWKWD>, <https://doi.org/10.18255/1818-1015-2018-6-726-733>
7. Михайлова, А.В. (2023). Анализ креативных сегментов в регионах Российской Федерации. *Экономика и предпринимательство*, 1 (150), 667–673. <https://www.elibrary.ru/LPZYPI>, <https://doi.org/10.34925/EIP.2023.150.1.133>
8. Ноакк, Н.В., Знаменская, А.Н. (2015). Факторы и феномены формирования потребительского спроса на киноконтент (опыт теоретического и экспериментального исследования). *Национальные интересы: приоритеты и безопасность*, 11.22 (307), 28–38. <https://www.elibrary.ru/TWDMKX>
9. Ноакк, Н.В., Неволин, И.В., Татарников, А.С. (2012). Методика прогнозирования выручки от проката кинофильмов. *Финансовая аналитика: проблемы и решения*, 48 (138), 17–24. <https://www.elibrary.ru/NVRNWR>
10. Павлова, Д.В. (2021). Российский кинематограф как потенциальный инструмент «мягкой силы» государства. *Информационные войны*, 2 (58), 93–96. <https://www.elibrary.ru/PXOLIZ>



11. Посов, И.А., Допира В.Е. (2019). Поиск плагиата в кодах программ. *Наука настоящего и будущего*, (1), 83–86. <https://www.elibrary.ru/DHIUQB>
12. Потапенко, А.А. (2019). *Семантические векторные представления текста на основе вероятностного тематического моделирования* [дисс.: 05.13.17] (147 с.). Москва: Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук». <https://www.elibrary.ru/DNXEFS>
13. Татарников, А.С. (2016). *Прогнозирование кассовых сборов в кинопрокате на основе математического моделирования и анализа зрительских эмоций* [автореф. дисс.: 08.00.13] (22 с.). Москва: Центр. эконом.-мат. ин-т РАН (ЦЭМИ). <https://www.elibrary.ru/ZPZKTP>
14. Эмих, П.В. (2015). Национальный кинематограф как основное направление государственной культурной политики в сфере услуг. *Вестник Орловского государственного университета. Серия: Новые гуманитарные исследования*, 2 (43), 431–432. <https://www.elibrary.ru/ULYLMV>
15. Юсупова, Г.М. (2021). Кино и телевидение: механизмы поддержки в современной Европе. *Наука телевидения*, 17 (2), 275–294. <https://www.elibrary.ru/XIZSHL>, <https://doi.org/10.30628/1994-9529-2021-17.2-275-294>
16. Ясницкий, Л.Н., Белобородова, Н.О., Медведева, Е.Ю. (2017). Методика нейросетевого прогнозирования кассовых сборов кинофильмов. *Финансовая аналитика: проблемы и решения*, 10.4 (334), 449–463. <https://www.elibrary.ru/YJYKNJ>, <https://doi.org/10.24891/fa.10.4.449>
17. Abidi, S.M.R., Xu, Y., Ni, J., Wang, X., & Wu, Z. (2020). Popularity prediction of movies: from statistical modeling to machine learning techniques. *Multimedia Tools and Applications*, 79, 35583–35617. <https://doi.org/10.1007/s11042-019-08546-5>
18. Adecola, O.D., Maitanmi, S.O., Kasali, F.A., Omotunde, A., Akande, O., Avorinde, O., Ajayi, W., & Mensah, Y. (2021). Movie Success Prediction Using Data Mining. *British Journal of Computer, Networking and Information Technology*, 4 (2), 22–30. <https://doi.org/10.52589/BJCNIT-CQOCIREC>
19. Chakraborty, P., Rahman, Md. Z., & Rahman, S. (2019). Movie Success Prediction using Historical and Current Data Mining. *International Journal of Computer Applications*, 178 (47), 1–5. <https://doi.org/10.5120/ijca2019919415>
20. Chen, Y., & Dai, Z. (2022). Mining of Movie Box Office and Movie Review Topics Using Social Network Big Data. *Frontiers in Psychology*, 13, 903380. <https://doi.org/10.3389/fpsyg.2022.903380>
21. Feng, K., & Liu, X. (2020). Adaptive Attention with Consumer Sentinel for Movie Box Office Prediction. *Complexity*, 6689304. <https://doi.org/10.1155/2020/6689304>
22. Gürbüz, A., Biçer, E., & Kaya, T. (2022). Prediction of Gross Movie Revenue in the Turkish Box Office Using Machine Learning Techniques. In: C. Kahraman, A.C. Tolga, S. Cevik Onar, S. Cebi, B. Oztaysi, & I.U. Sari (eds.), *Intelligent and Fuzzy Systems INFUS*

2022. *Lecture Notes in Networks and Systems* (vol. 505, pp. 86—92). Springer, Cham. [https://doi.org/10.1007/978-3-031-09176-6\\_10](https://doi.org/10.1007/978-3-031-09176-6_10)

23. Li, D., & Liu, Z-P. (2022). Predicting Box-Office Markets with Machine Learning Methods. *Entropy*, 24 (5), 711. <https://doi.org/10.3390/e24050711>

24. Lu, W. (2019). Research on Movie Box Office Prediction Model with AHP Method. In: *IMMS '19: Proceedings of the 2nd International Conference on Information Management and Management Sciences* (pp. 177—181). New York: Association for Computing Machinery. <https://doi.org/10.1145/3357292.3357322>

25. Madongo, C.T., & Zhongjun, T. (2023). A movie box office revenue prediction model based on deep multimodal features. *Multimedia Tools and Applications*, 82, 31981—32009. <https://doi.org/10.1007/s11042-023-14456-4>

26. Meenakshi, K., Maragatham, G., Agarwal, N., & Ishitha, G. (2018). A Data mining Technique for Analyzing and Predicting the success of Movie. *Journal of Physics: Conference Series*, 1000, 012100. <https://doi.org/10.1088/1742-6596/1000/1/012100>

27. Murschetz, P.C., Bruneel, C., Guy, J.-L., Haughton, D., Lemerrier, N., McLaughlin, M.-D., Mentzer, K., Vialle, Q., Zhang, C., Murschetz, P. C., & Bakhtawar, B. (2020). Movie Industry Economics: How Data Analytics Can Help Predict Movies' Financial Success. *Nordic Journal of Media Management*, 1 (3), 339—359. <https://doi.org/10.5278/njmm.2597-0445.5871>

28. Ni, Y., Dong, F., Zou, M., & Li, W. (2022). Movie Box Office Prediction Based on Multi-Model Ensembles. *Information*, 13 (6), 299. <https://doi.org/10.3390/info13060299>

29. Sivakumar, P., Rajeswaren, V., Abishankar, K., Ekanayake, J., & Mehendran, Y. (2021). *Movie Success and Rating Prediction Using Data Mining Algorithms*. <https://doi.org/10.13140/RG.2.2.14697.42085>

30. Yoo, B.-K., & Kim, S.-H. (2023). Movie Box Office Prediction at the Distribution Stage Using Text Mining of Movie Reviews. *Korean Logistics Research Association*, 33 (1), 95—105. <https://doi.org/10.17825/klr.2023.33.1.95>

31. Zhao, J., Xiong, F., & Jin, P. (2022). Enhancing Short-Term Sales Prediction with Microblogs: A Case Study of the Movie Box Office. *Future Internet*, 14 (5), 141. <https://doi.org/10.3390/fi14050141>

32. Zhou, Y., & Yen, G.G. (2018). Evolving Deep Neural Networks for Movie Box-Office Revenues Prediction. In: *2018 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1—8). Rio de Janeiro. <https://doi.org/10.1109/CEC.2018.8477691>

33. Zhou, Y., Zhang, L., & Yi, Z. (2019). Predicting movie box-office revenues using deep neural networks. *Neural Computing and Applications*, 31, 1855—1865. <https://doi.org/10.1007/s00521-017-3162-x>

## REFERENCES

1. Abidi, S.M.R., Xu, Y., Ni, J., Wang, X., & Wu, Z. (2020). Popularity prediction of movies: From statistical modeling to machine learning techniques. *Multimedia Tools and Applications*, 79, 35583–35617. <https://doi.org/10.1007/s11042-019-08546-5>
2. Adekola, O.D., Maitanmi, S.O., Kasali, F.A., Omotunde, A., Akande, O., Avorinde, O., Ajayi, W., & Mensah, Y. (2021). Movie Success Prediction Using Data Mining. *British Journal of Computer, Networking and Information Technology*, 4 (2), 22–30. <https://doi.org/10.52589/bjcnit-cqocirec>
3. Astafyev, S.A., Khomkalov, G.V., & Tolstoukhova, I.S. (2021). Kreativnaya ekonomika kak klyuchevoy element ustoychivogo razvitiya territoriy [Creative economy as a key element of sustainable territorial development]. *Baikal Research Journal*, 12, (3), 16. (In Russ.)
4. Chakraborty, P., Rahman, Md. Z., & Rahman, S. (2019). Movie success prediction using historical and current data mining. *International Journal of Computer Applications*. 178, (47), 1–5. <https://doi.org/10.5120/ijca2019919415>
5. Chen Y., & Dai Z. (2022). Mining of movie box office and movie review topics using social network big data. *Frontiers in Psychology*, 13, Article 903380. <https://doi.org/10.3389/fpsyg.2022.903380>
6. Dobrotvorskaya, D.A. (2023) Rol' gosudarstvennoy podderzhki v razvitiy kreativnykh industriy v usloviyakh sanktsiy i ogranicheniy [The role of state support in the development of creative industries under conditions of sanctions and restrictions]. In I.M. Stepnov, J.A. Kovalchuk (Eds.), *Upravlenie aktivami—2023: Biznes-modeli v epokhu izmeneniya delovogo klimata* [Asset management 2023: Business models in an era of changing business climate]. Moscow: Federal State Budgetary Institution of Science Institute of Market Problems of the Russian Academy of Sciences, 297–301. (In Russ.)
7. Dozhdikov, A.V. (2023). Prognozirovaniye rezul'tatov kinoprokata s pomoshch'yu mashinnogo obucheniya [Predicting film distribution results using machine learning]. *Voprosy Teoreticheskoy Ekonomiki*, (4), 93–114. (In Russ.)
8. Dozhdikov, A.V. (2024). Opredeleniye investitsionnogo uspekha i ego faktorov dlya rossiyskogo kino v prokate s pomoshch'yu mashinnogo obucheniya [Determination of investment success and its factors for Russian cinema at the box office using machine learning]. *Finansy: Teoriya i Praktika*, 28, (1), 188–203. (In Russ.)
9. Emich, P.V. (2015). Natsional'nyy kinematograf kak osnovnoe napravleniye gosudarstvennoy kul'turnoy politiki v sfere uslug [National cinema as the primary focus of state cultural policy in the service sector]. *Vestnik Orlovskogo Gosudarstvennogo Universiteta. Seriya: Novye Gumanitarnye Issledovaniya*, (2), 112–114. (In Russ.)
10. Feng, K., & Liu, X. (2020). Adaptive attention with consumer sentiment for movie box office prediction. *Complexity*, Article 6689304. <https://doi.org/10.1155/2020/6689304>
11. Gorshkov, M.K., & Sheregi, F.E (2019). *Rossiyskaya molodezh' v kontekste sotsiologicheskogo analiza* [Russian youth within the context of sociological analysis].

Moscow: Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences. (In Russ.)

12. Gürbüz, A., Biçer, E., & Kaya, T. (2022). Prediction of gross movie revenue in the Turkish box office using machine learning techniques. In C. Kahraman, A.C. Tolga, S.Ç. Onar, S. Cebi, B. Oztaysi, & I.U. Sari (Eds.), *Intelligent and Fuzzy Systems. INFUS 2022. Lecture Notes in Networks and Systems* (Vol. 505). Springer, Cham. [https://doi.org/10.1007/978-3-031-09176-6\\_10](https://doi.org/10.1007/978-3-031-09176-6_10)

13. Karyaeva, M.S., Braslavski, P.I., & Sokolov, V.A. (2018). Vektornoe predstavlenie slov s semanticheskimi otnosheniyami: Eksperimental'nye nablyudeniya [Word embedding for semantically relative words: An experimental study]. *Modelirovanie i Analiz Informatsionnykh Sistem*, 25, (6), 726–733. (In Russ.)

14. Li, D., & Liu, Z.-P. (2022). Predicting box-office markets with machine learning methods. *Entropy*, 24 (5), 711. <https://doi.org/10.3390/e24050711>

15. Lu, W. (2019). Research on movie box office prediction model with AHP method. In *Proceedings of the 2nd International Conference on Information Management and Management Sciences (IMMS '19)* (pp. 177–181). Association for Computing Machinery, New York. <https://doi.org/10.1145/3357292.3357322>

16. Madongo, C.T., & Zhongjun, T. A. (2023). A movie box office revenue prediction model based on deep multimodal features. *Multimedia Tools and Applications*, 82, 31981–32009. <https://doi.org/10.1007/s11042-023-14456-4>

17. Meenakshi K., Maragatham G., Agarwal N., & Ishitha G. (2018). A data mining technique for analyzing and predicting the success of movie. *Journal of Physics: Conference Series*, 1000, Article 012100. <https://doi.org/10.1088/1742-6596/1000/1/012100>

18. Mikhaylova, A.V. (2023). Analiz kreativnykh segmentov v regionakh Rossiyskoy Federatsii [Analysis of creative segments in the regions of the Russian Federation]. *Ekonomika i Predprinimatel'stvo*, 1, 667–673. (In Russ.)

19. Murschetz, P. C., Bruneel, C., Guy, J.-L., Haughton, D., Lemerrier, N., McLaughlin, M.-D., Mentzer, K., Vialle, Q., Zhang, C., Murschetz, P. C., & Bakhtawar, B. (2020). Movie industry economics: How data analytics can help predict movies' financial success. *Nordic Journal of Media Management*, 1 (3), 339–359. <https://doi.org/10.5278/njmm.2597-0445.5871>

20. Ni, Y., Dong, F., Zou, M., & Li, W. (2022). Movie box office prediction based on multi-model ensembles. *Information*, 13 (6), 299. <https://doi.org/10.3390/info13060299>

21. Noakk, N.V., & Znamenskaya, A.N. (2015). Faktory i fenomeny formirovaniya potrebitel'skogo sprosa na kinokontent (opyt teoreticheskogo i eksperimental'nogo issledovaniya) [Factors and phenomena of generating consumer demand for cinematographic content (theoretical and experimental studies)]. *Natsional'nye Interesy: Priority i Bezopasnost'*, 11 (22), 28–38. (In Russ.)

22. Noakk, N.V., Nevolin, I.V., & Tatarnikov, A.S. (2012). Metodika prognozirovaniya vyruchki ot prokata kinofil'mov [A methodology for predicting revenue from movie rentals]. *Finansovaya Analitika: Problemy i Resheniya*, (48), 17–24. (In Russ.)

23. Pavlova, D.V. (2021). Rossiyskiy kinematograf kak potentsial'nyy instrument "myagkoy sily" gosudarstva [Russian cinema as a potential soft power tool of the government]. *Informatsionnye Voyny*, (2), 93–96. (In Russ.)
24. Posov, I.A., & Dopira V.E. (2019). Poisk plagiata v kodakh program [Search for plagiarism in program codes. Science of the present and future]. *Nauka Nastoyashchego i Budushchego*, 1, 83–86. (In Russ.)
25. Potapenko, A.A. (2019). *Semanticheskie vektornye predstavleniya teksta na osnove veroyatnostnogo tematicheskogo modelirovaniya* [Semantic vector representations of text based on probabilistic topic modeling] [PhD dissertation]. (In Russ.)
26. Sivakumar, P., Rajeswaren, V., Abishankar, K., Ekanayake, J., & Mehendran, Y. (2021). Movie success and rating prediction using data mining algorithms. In *International Research Conference of Uva Wellassa University (IRCUWU-2020)*. [https://doi.org/10.13140/RG.2.2.14697.42085\\_](https://doi.org/10.13140/RG.2.2.14697.42085_)
27. Tatarnikov, A.S. (2016). *Prognozirovaniye kassovykh sborov v kinoprote na osnove matematicheskogo modelirovaniya i analiza zritel'skikh emotsiy* [Forecasting box office revenue in cinema based on mathematical modeling and analysis of viewer emotions] [PhD thesis]. (In Russ.)
28. Yasnitskii, L.N., Beloborodova, N.O., & Medvedeva, E.Yu. (2017). Metodika neyrosetevogo prognozirovaniya kassovykh sborov kinofil'mov [The method of neural network forecasting of box-office grosses of movies]. *Finansovaya Analitika: Problemy i Resheniya*, 10 (4), 449–463. (In Russ.)
29. Yoo, B.-K., & Kim, S.-H. (2023). Movie box office prediction at the distribution stage using text mining of movie reviews. *Korean Logistics Research Association*, 33 (1), 95–105. Retrieved March 25, 2024, from [https://www.researchgate.net/publication/369121771\\_Movie\\_Box\\_Office\\_Prediction\\_at\\_the\\_Distribution\\_Stage\\_Using\\_Text\\_Mining\\_of\\_Movie\\_Reviews](https://www.researchgate.net/publication/369121771_Movie_Box_Office_Prediction_at_the_Distribution_Stage_Using_Text_Mining_of_Movie_Reviews)
30. Yusupova, G.M. (2021). Kino i televidenie: Mekhanizmy podderzhki v sovremennoy Evrope [Film and television: Financial support facilities in modern Europe]. *Nauka Televideniya—The Art and Science of Television*, 17, (2), 275–294. (In Russ.)
31. Zhao, J., Xiong, F., & Jin, P. (2022). Enhancing short-term sales prediction with microblogs: A case study of the movie box office. *Future Internet*, 14 (5), 141. <https://doi.org/10.3390/fi14050141>
32. Zhou, Y., & Yen, G.G. (2018). Evolving deep neural networks for movie box-office revenues prediction. In *IEEE Congress on Evolutionary Computation (CEC)* (pp. 1–8). Rio de Janeiro, Brazil.
33. Zhou, Y., Zhang, L., & Yi, Z. (2019). Predicting movie box-office revenues using deep neural networks. *Neural Computing and Applications*, 31, 1855–1865. <https://doi.org/10.1007/s00521-017-3162-x>

#### СВЕДЕНИЯ ОБ АВТОРЕ

##### **ДОЖДИКОВ АНТОН ВАЛЕНТИНОВИЧ**

кандидат политических наук,  
старший научный сотрудник кафедры ЮНЕСКО по социальным  
и гуманитарным наукам,  
Институт социально-политических исследований  
Федерального научно-исследовательского социологического  
центра Российской академии наук  
11933, Россия, г. Москва, ул. Фотиевой, д. 6, к. 1

**ResearcherID: KYP-9166-2024**

**ORCID: 0000-0002-1069-1648**

**e-mail: antondnn@yandex.ru**

#### ABOUT THE AUTHOR

##### **DODZHNIKOV ANTON VALENTINOVICH**

Cand. Sci. (Political Science),  
Senior Research Fellow at the UNESCO Department of Social and  
Humanitarian Sciences,  
Institute of Socio-Political Research,  
Federal Center of Theoretical and Applied Sociology of the Russian  
Academy of Sciences,  
6, korp. 1, Fotiyevoy, Moscow 11933, Russia

**ResearcherID: KYP-9166-2024**

**ORCID: 0000-0002-1069-1648**

**e-mail: antondnn@yandex.ru**