|  |
| --- |
| Пермский филиал федерального государственного автономного образовательного учреждения высшего образования  «Национальный исследовательский университет  «Высшая школа экономики»  Факультет экономики, менеджмента и бизнес-информатики |
|  |
| Чепоков Елизар Сергеевич  **Разработка MVP системы машинного обучения для прогнозирования рентабельности кинобизнеса**  *Выпускная квалификационная работа*  студента образовательной программы «Программная инженерия» по направлению подготовки 09.03.04 Программная инженерия   |  |  | | --- | --- | | Рецензент  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | Руководитель  Профессор кафедры информационных  технологий в бизнесе,  Доктор технических наук \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Л. Н. Ясницкий | |

Пермь, 2022 год

Аннотация

Название: Разработка MVP системы машинного обучения для прогнозирования рентабельности кинобизнеса.

Автор: Чепоков Елизар Сергеевич, студент четвертого курса образовательной программы «Программная инженерия».

Руководитель: Ясницкий Леонид Нахимович, профессор кафедры информационных технологий в бизнесе, доктор технических наук.

Данная выпускная квалификационная работа посвящена разработке системы машинного обучения для прогнозирования рентабельности кинобизнеса. В первой главе проанализированы методы прогнозирования и выявлены взаимосвязи рентабельности кинобизнеса, а также выдвинуты требования, предъявляемые системе. Вторая глава описывает процесс проектирования алгоритмической части системы и анализ данных для машинного обучения. Третья глава посвящена реализации конечной системы. В заключении описаны итоги выполненной работы.

Работа включает \_\_ страниц формата А4, из них в основной части \_\_ страниц.

Основная часть работы включает в себя \_\_ иллюстрацию и \_\_ таблиц.

Библиографический список состоит из \_\_ публикаций.

Работа включает в себя \_\_ приложений.

# Оглавление

[Аннотация 2](#_Toc90153680)

[Оглавление 3](#_Toc90153681)

[Введение 4](#_Toc90153682)

[Глава 1 Анализ 8](#_Toc90153683)

[1.1 Анализ особенностей кинобизнеса 8](#_Toc90153684)

[1.2 Анализ особенностей систем машинного обучения 8](#_Toc90153685)

[1.3 Обзор существующих методов прогнозирования 8](#_Toc90153686)

[1.4 Постановка требований к системе 8](#_Toc90153687)

[1.5 Выводы по первой главе 8](#_Toc90153688)

[Глава 2 Проектирование 9](#_Toc90153689)

[2.1 Проектирование модуля сбора и систематизации данных 9](#_Toc90153690)

[2.2 Проектирование системы машинного обучения 9](#_Toc90153691)

[2.3 Проектирование графического интерфейса 9](#_Toc90153692)

[2.4 Выводы по второй главе 9](#_Toc90153693)

[Глава 3 Разработка и тестирование 10](#_Toc90153694)

[3.1 Обоснование выбора средств разработки 10](#_Toc90153695)

[3.2 Реализация модуля сбора и систематизации данных 10](#_Toc90153696)

[3.3 Реализация системы машинного обучения 10](#_Toc90153697)

[3.4 Реализация графического интерфейса 10](#_Toc90153698)

[3.5 Выводы по третьей главе 10](#_Toc90153699)

[Заключение 11](#_Toc90153700)

[Список сокращений и условных обозначений 12](#_Toc90153701)

[Библиографический список 13](#_Toc90153702)

[ПРИЛОЖЕНИЕ А 14](#_Toc90153703)

Введение

Киноиндустрия вносит огромный вклад как в общую социальную жизнь человека, так и в развитие экономики стран, она стимулирует многочисленные внешние эффекты: потребление определенных товаров, посещение представленных в кинокартинах стран, развитие индустрии игр и т.п. По данным отчета «Motion picture association» Американской ассоциации кинокомпаний, за 2019 год суммарные кассовые сборы всех вышедших в мировой кинотеатральный прокат фильмов составили $98,3 млрд., что на 6,8% больше, чем в 2018 году. Однако в 2020 году из-за сложившейся эпидемиологической ситуации во всем мире кассовые сборы резко упали и составили $80,8 млрд., на 17.8% меньше чем в 2019 году, количество проданных билетов составило 223 млн., против 1,3 млрд в 2019 году. Так же стоит обратить внимание, что доля кинотеатрального проката с каждым годом становится меньше, когда доля онлайн просмотра на стриминговых сервисах растет, так, в 2020 году стриминговые сервисы собрали $61,8 млрд (76% от суммарных кассовых сборов), что на $14,6 млрд больше чем в 2019 году ($47,2 млрд). На долю США и Канады приходится 40% всех кассовых сборов ($32,2 млрд). Россия же находится на 8-ом месте по кассовым сборам, суммарные кинотеатральные сборы в 2020 году в России составили лишь $192 млн.

Производство кинофильмов, было и остается одним из наиболее рискованных видов предпринимательства, особенно во времена ограничений на посещение общественных мест. Затрачивая огромные бюджеты в несколько сотен миллионов долларов на производство фильма кинокомпании ожидают соответствующие доходы от показа фильма в кинотеатрах и продаж на стриминговых сервисах, но инвесторы отмечают, что предсказать финансовый успех фильма крайне сложно даже для опытных участников рынка. По данным сайта «The numbers» лишь около 30% фильмов окупает затраты на их производство и приносит прибыль, так как кинокартина считается прибыльной только при превышении кассовых сборов больше чем в два раза поставленного бюджета. Расширение знаний относительно факторов успеха фильмов позволит уменьшить долю неудачных инвестиций в кинопроизводство. Предварительная оценка будущего коммерческого успеха кинофильмов может позволить инвесторам сузить круг проектов, требующих более тщательного последующего анализа их инвестиционной привлекательности.

Актуальность данной выпускной квалификационной работы заключается в том, что выпуск кинокартины в наши дни является большим риском, не только для киностудий и режиссеров, но и для инвесторов, система прогнозирования способна снизить риски неудачных инвестиций в кинопроизводство на ранних этапах создания кинокартины, а также спрогнозировать целесообразность выпуска кинокартины на экраны кинотеатров с сопутствующими затратами на рекламу и аренду кинозалов. В настоящее время решения о целесообразности реализации кинопроектов принимаются в большинстве случаев исключительно на основе интуиции и предыдущего опыта продюсерских организаций, что способствует неоправданному росту рисков инвесторов, вкладывающих значительные финансовые средства в кинопроизводство. Хотя, интуиция и опыт являются важнейшими профессиональными качествами продюсера, для роста обоснованности принимаемых решений необходимо, чтобы они основывались на комплексном анализе сильных и слабых сторон кинопроекта до его запуска в производство.

Таким образом, актуальность исследования основывается на необходимости решения важных задач, таких как, научное обоснование и создание эффективно функционирующего механизма прогнозирования коммерческого потенциала кинопроекта, принятия рациональных управленческих решений о целесообразности его реализации, определение направлений и принципов эффективного управления кинематографическим бизнес-процессом, в частности, на ранних стадиях создания кинофильмов.

***Объектом*** исследования в данной работе являются система алгоритм машинного обучения. ***Предметом*** исследования выступает разработка алгоритма системы машинного обучения.

***Целью*** данной работы является разработка системы машинного обучения для прогнозирования рентабельности кинобизнеса.

Задачи для достижения поставленной цели:

1. Провести анализ систем машинного обучения и методов прогнозирования.
   1. Провести анализ особенностей кинобизнеса.
   2. Провести анализ особенностей построения систем машинного обучения.
   3. Формализовать существующие методы прогнозирования рентабельности кинобизнеса.
   4. Конкретизировать требования к разрабатываемой системе.
2. Спроектировать систему.
   1. Спроектировать модуль сбора и систематизации данных.
   2. Спроектировать системы машинного обучения.
   3. Спроектировать графический интерфейс разрабатываемой системы.
3. Реализовать систему машинного обучения для прогнозирования рентабельности кинобизнеса.
   1. Обосновать выбор технологий для разработки.
   2. Реализовать модуль сбора и систематизации данных.
   3. Реализовать систему машинного обучения.
   4. Реализовать графический интерфейс разрабатываемой системы.
   5. Провести тестирование системы.

При проектировании и разработке системы машинного обучения для прогнозирования рентабельности кинобизнеса были использованы такие общенаучные методы исследования, как анализ и синтез, системный подход, статистический анализ, методы сравнений, методы объектно-ориентированного проектирования и объектно-ориентированного программирования.

Одними из первых исследователей, применивших в кинобизнесе метод экономико-математического моделирования, были J.Prag и J.Casavant, которые в 1994 опубликовали статью [6] с сообщением о создании регрессионной модели на основе выборки из 625 американских фильмов. Рассматриваемый ими набор входных переменных включал производственный бюджет, критические обзоры, наличие звезд, наличие франшизы, наличие премий, жанр и рейтинг. Аппарат нейронных сетей для прогнозирования кассовых сборов фильмов был впервые применен в 2002 году американскими учеными R.Sharda и D.Delen. В 2006 году эти же авторы [7] построили модели на основе логистической регрессии, дискриминантного анализа, классификационного и регрессионного дерева, а также нейронной сети, показавшей наилучший результат.

Практическая значимость данной работы заключается в возможности применения разрабатываемой системы для представления методических рекомендаций по рационализации продюсерской деятельности, обеспечивающих достижение высоких экономических результатов кино-отрасли и отдельных кинокартин.

Научная новизна работы состоит в разработке состоит в разработке моделей, подходов и методик, а также системы, способствующих анализу коммерческого успеха кинофильмов, позволяющего сделать количественно обоснованные выводы об их коммерческом потенциале.

1. Анализ

В данной главе будут рассмотрены особенности кинобизнеса и существующих методов прогнозирования. Будут выбраны критерии отбора данных для последующего анализа системами машинного обучения, выявления их преимуществ, недостатков и особенностей. Результатом первой главы являются поставленные требования к разрабатываемой системе.

* 1. Анализ особенностей кинобизнеса

Описание критериев «популярности кинокартин»

Приведение и анализ кассовых сборов кинокартин

* 1. Анализ особенностей систем машинного обучения

Варианты обучения

Варианты сбора данных

* 1. Обзор существующих методов прогнозирования

Анализ исследований (J.Prag и J.Casavant) и (R.Sharda и D.Delen)

* 1. Постановка требований к системе

Конкретизация требований

* 1. Выводы по первой главе

В результате анализа особенностей кинобизнеса и существующих методов прогнозирования были выбраны критерии отбора данных для последующего анализа системами машинного обучения. Были выдвинуты функциональные требования, предъявляемые к системе. Проведен краткий обзор и сравнение систем и способов написания нейронных сетей. Рассмотрены плюсы и минусы аналогов и определены преимущества разрабатываемой системы.

1. Проектирование

В представленной главе спроектированы и рассмотрены методы и алгоритмы, которые потребуются для реализации системы и для удовлетворения требований, выявленных в ходе анализа. Результатом данной главы является спроектированная система и её интерфейс.

* 1. Проектирование модуля сбора и систематизации данных
  2. Проектирование системы машинного обучения
  3. Проектирование графического интерфейса
  4. Выводы по второй главе

На основании результатов, полученных из первой главы, была спроектирована алгоритмическая часть системы.

1. Разработка и тестирование

Данная глава обозревает реализацию и тестирование конечного продукта. В данной главе рассматриваются все основы создания системы и ее тестирование. Итоговым результатом третьей главы является конечный продукт.

* 1. Обоснование выбора средств разработки
  2. Реализация модуля сбора и систематизации данных
  3. Реализация системы машинного обучения
  4. Реализация графического интерфейса
  5. Выводы по третьей главе

Заключение

Список сокращений и условных обозначений

MVP (minimum viable product) – Минимально жизнеспособный продукт, продукт обладающий минимальными, но достаточными для удовлетворения первых потребителей функциями.

UML (Unified Modeling Language) – унифицированный язык моделирования, система обозначений, которую можно применять для объектно-ориентированного анализа и проектирования

UI (user interface) – Интерфеейс пользователя, он же пользовательский интерфейс, интерфейс, обеспечивающий передачу информации между пользователем-человеком и программно-аппаратными компонентами компьютерной системы

ПО – программное обеспечение.

Библиографический список

1. Черепанов Ф.М., Ясницкий Л.Н. Симулятор нейронных сетей «Нейросимулятор 1.0». // Свидетельство об отраслевой регистрации разработки № 8756. Зарегистрировано в Отраслевом фонде алгоритмов и программ 12.07.2007.
2. Ясницкий Л.Н. Введение в искусственный интеллект: учеб. пособие для студ. высш. учеб. заведений/ Л.Н. Ясницкий. – 3-е изд. – М.: Издательский центр «Академия», 2010. – 176с.
3. Ясницкий Л.Н. Пермская научная школа искусственного интеллекта и ее инновационные проекты / Л.Н.Ясницкий, В.В.Бондарь, С.Н.Бурдин и др.; под ред. Л.Н.Ясницкого. – 2-е изд. – Москва-Ижевск: НИЦ «Регулярная и хаотическая динамика», 2008. – 75 с.
4. Internet Movie Data Base – [Электронный ресурс]/ Режим доступа: <http://imdb.com>
5. Kinopoisk – [Электронный ресурс]/ Режим доступа: <http://kinopoisk.ru>
6. Prag J. and Casavant J. An Empirical Study of the Determinants of Revenue and Marketing Expenditures in the Motion Picture Industry // Journal of Cultural Economics. – 1994. – No 18(3). – Pp. 217-235.
7. Sharda R., Delen D. Predicting box-office success of motion pictures with neural networks // Expert Systems with Applications. – 2006. – No 30. – Pp. 243-254.

ПРИЛОЖЕНИЕ А