



Факультет социально-экономических и
компьютерных наук

Программная инженерия

Пермь, 2022

РАЗРАБОТКА MVP СИСТЕМЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ РЕНТАБЕЛЬНОСТИ КИНОБИЗНЕСА

Работу выполнил: студент группы ПИ-18-2, 4-го
курса факультета социально-экономических и
компьютерных наук:
Чепокоев Е. С.

Научный руководитель: профессор кафедры
информационных технологий в бизнесе, доктор
технических наук:
Ясницкий Л. Н.



1. [Основная информация](#)
2. [Постановка цели и задач](#)
3. [Анализ предметной области](#)
4. [Обзор аналогов](#)
5. [Постановка требований](#)
6. [Проектирование поведения системы](#)
7. [Проектирование алгоритма сбора данных](#)
8. [Проектирование нейронной сети](#)
9. [Обоснование средств разработки](#)
10. [Реализация модуля сбора данных](#)
11. [Реализация нейронной сети](#)
12. [Реализация интерфейса](#)
13. [Заключение](#)

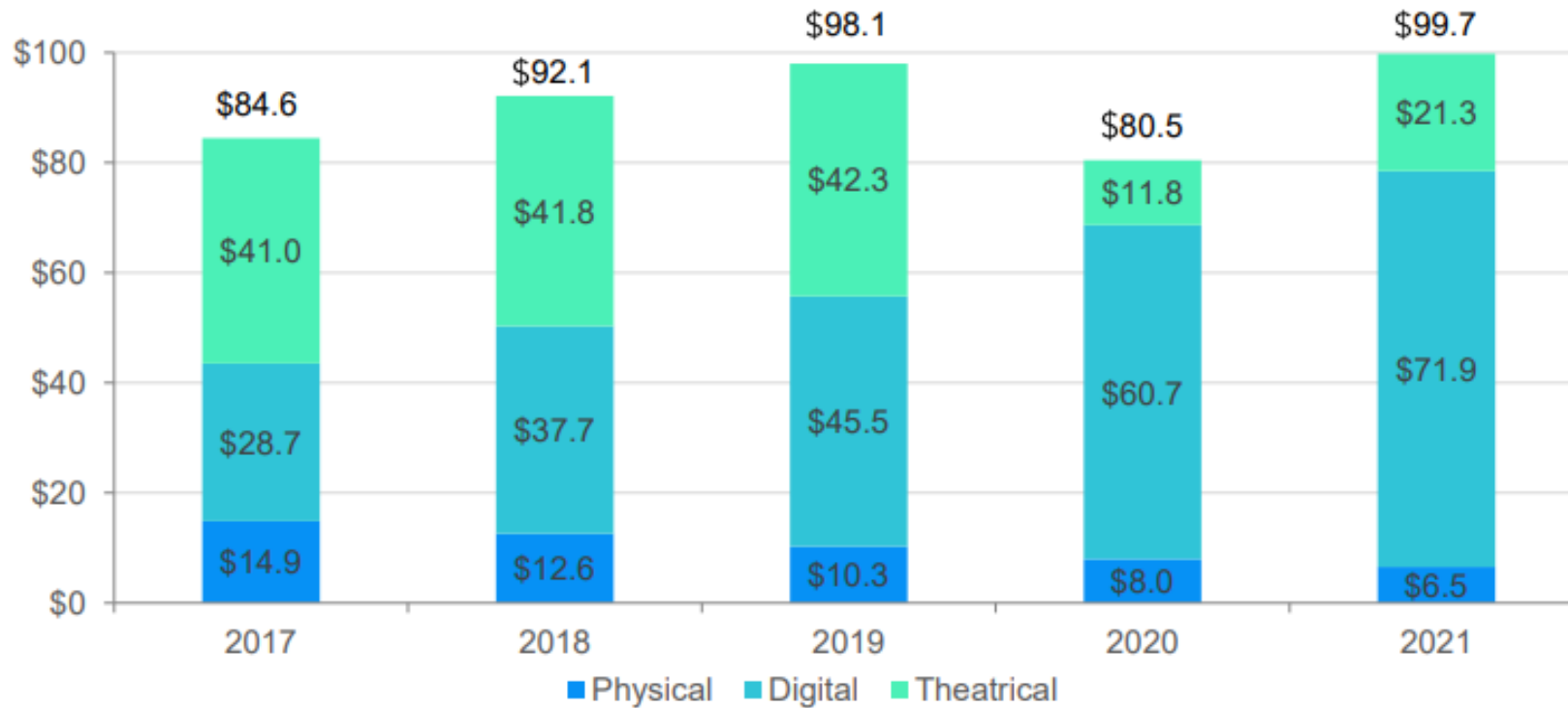


Рисунок 1 – Суммарная прибыль мирового рынка кинопроката

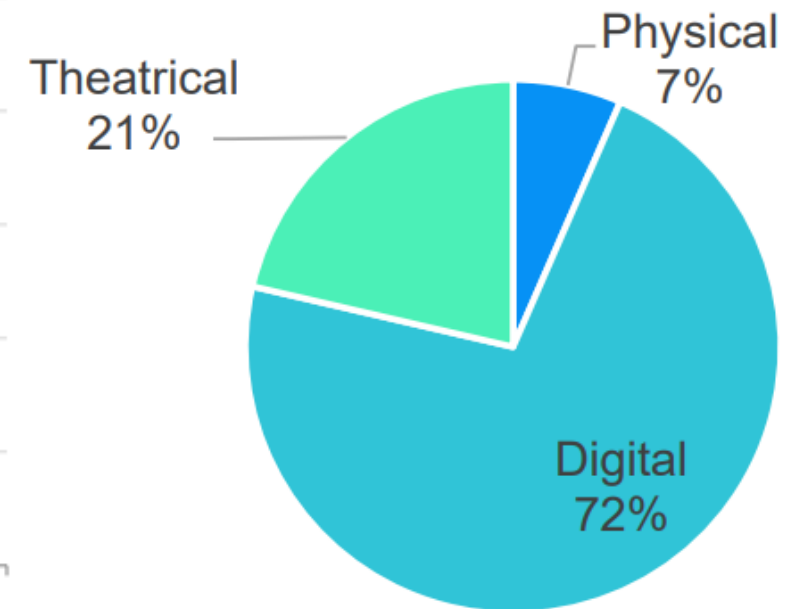


Рисунок 2 – Доли Театрального, Цифрового и Физического кинопроката

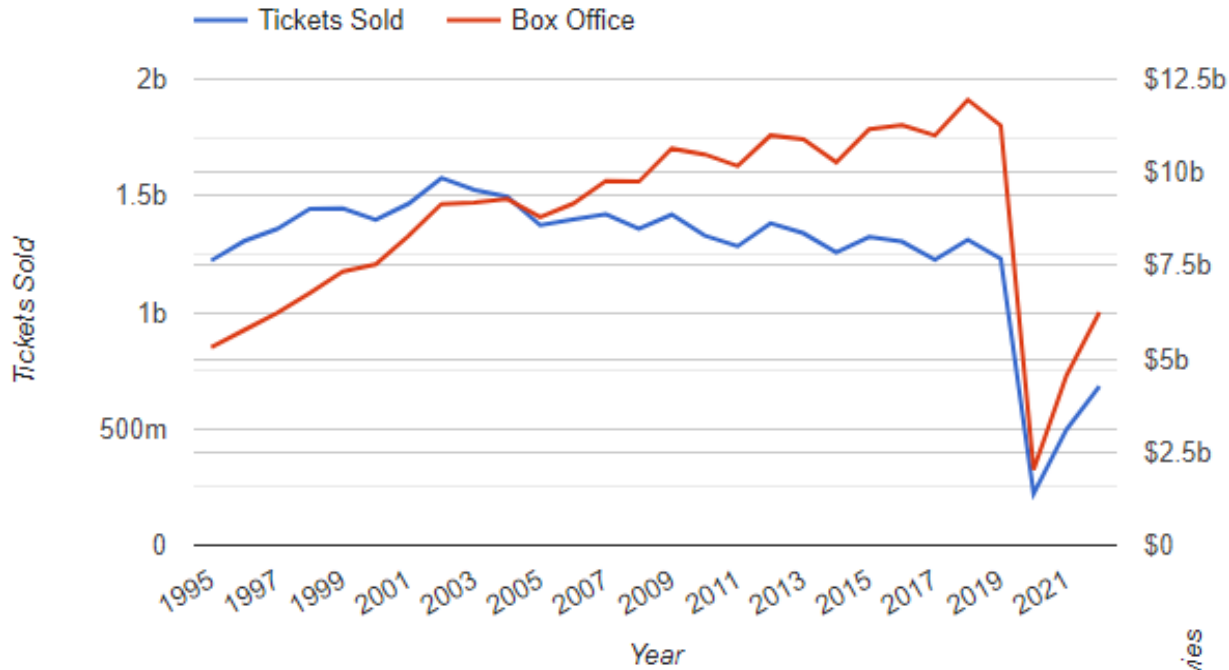


Рисунок 3 – Ежегодные продажи билетов

Таблица 1 – Итоги продаж последних 3-х лет

Год	Общее количество проданных билетов	Итоговые кассовые сборы	Общее количество выпущенных фильмов
2021	497,910,076	\$4,565,836,737	60 (35)
2020	221,762,724	\$2,033,566,047	34 (23)
2019	1,228,766,802	\$11,255,506,522	87 (44)

Released Movies

Box Office (USD)

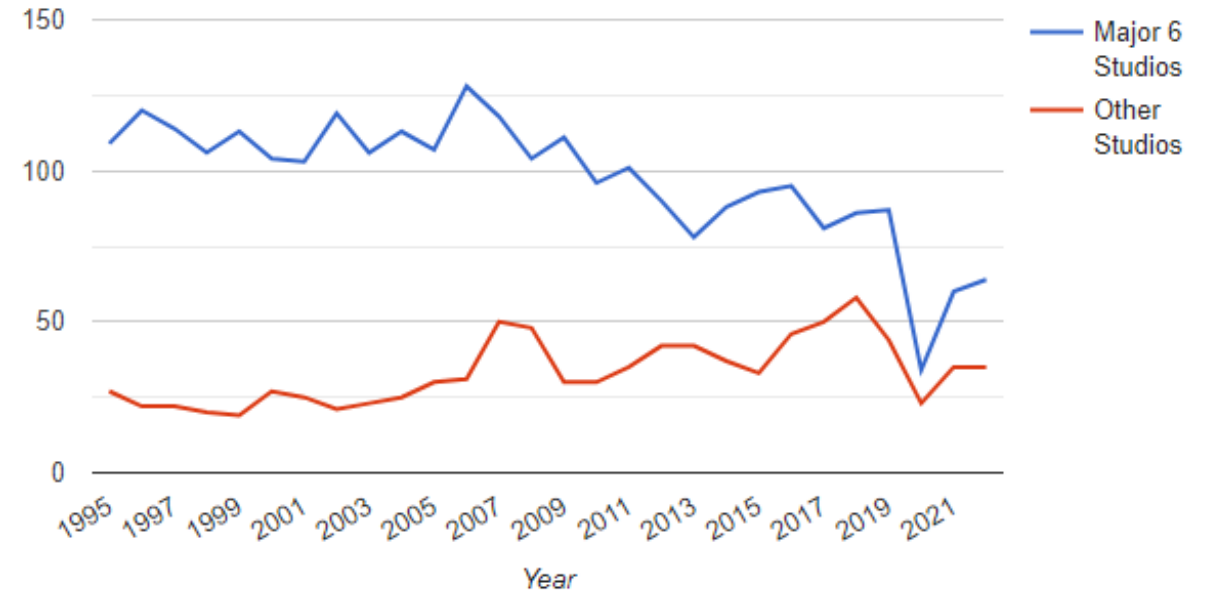


Рисунок 4 – Количество ежегодно выпускаемых картин

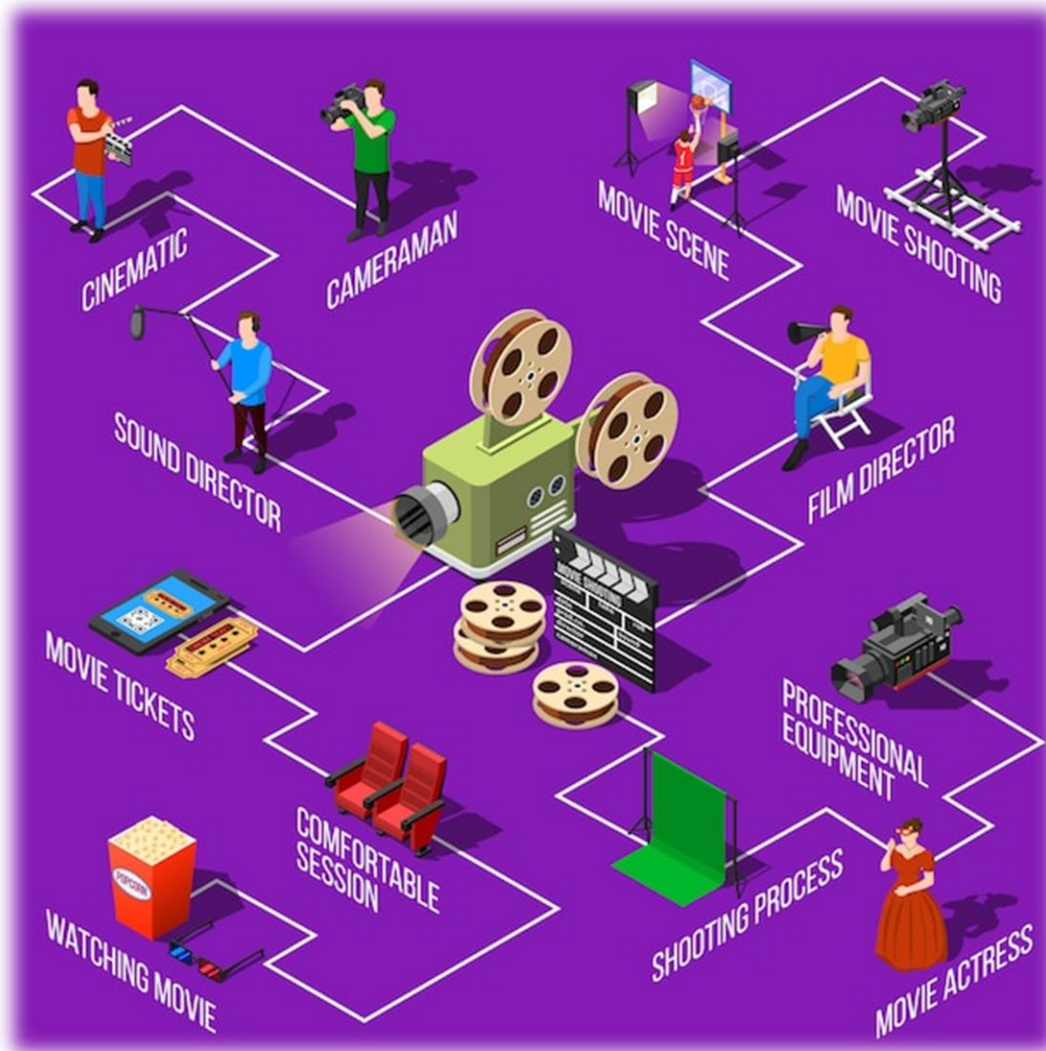
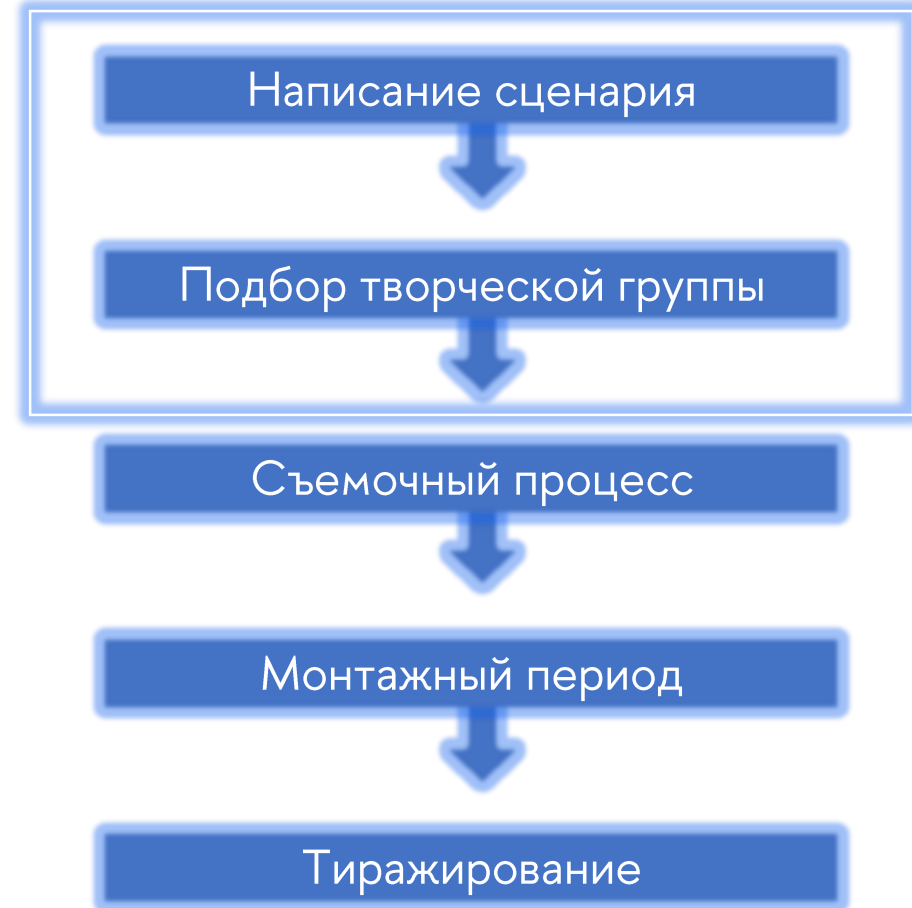


Рисунок 5 – Процесс создания кинокартин

Снижение производственных и инвестиционных рисков





Объект: Алгоритмы систем машинного обучения.

Предмет: Разработка алгоритма системы машинного обучения.

Цель: Разработать систему машинного обучения для прогнозирования рентабельности кинобизнеса.

Задачи:

1. Провести анализ области машинного обучения и методов прогнозирования.
2. Спроектировать систему.
3. Реализовать и протестировать систему.



Таблица 2 – Выбранные нейронные сети

Название	Сферы применения	Сильные стороны	Слабые стороны
<i>Персептрон</i>	Прогнозирование, управление агентами, слабая возможность классификации	Способность к обучению по простому и эффективному алгоритму	Качество прогноза и точность построенной модели зависит от числа знаний, используемых при построении модели
<i>Нейронная сеть Джордана</i>	«То же что и в Персептроне»	Нейроны имеют обратную связь	–
<i>Нейронная сеть Коско</i>	Выявление ассоциаций	Адаптивность	Малый объём памяти
<i>Машина Гельмгольца</i>	Создание представлений, мимикрия	Анатомический, анализируемый с теорией информации	–



Год выпуска фильма	Количество премьерных показов;	Наличие главного героя, заметно превосходящего в силе стальных
Бюджет фильма	Количество наград у съемочной группы	Наличие спецэффектов
Страна производитель	Рейтинг актеров	Рейтинг с сайта «IMDb»
Доминирующий жанр фильма	Возраст режиссера на момент создания фильма	Количество положительных оценок с сайта «IMDb»
Длительность фильма	Рейтинг режиссера	Рейтинг пользователей с сайта «Metascore»
Возрастное ограничение	Количество внушительных наград у режиссера	Рейтинг с сайта «Rotten Tomatoes»
Произведение, ставшее основой сценария	Количество внушительных наград у актеров	Рейтинг критиков с сайта «Rotten Tomatoes»
Наличие франшизы	Сезон выхода	Оценка обзоров пользователей с сайта «Rotten Tomatoes»
Критические обзоры	Пол режиссера	Рейтинг пользователей с сайта «Rotten Tomatoes».
Наличие звезд	Наличие вымышленных персонажей	
Конкуренция в период релиза	Наличие в фильме злодея	
Качество съемки		



Таблица 3 – Набор входных параметров

№	Параметр	Изначальные значения	Кодировка
X1	Бюджет	100000 – 1000000000	1 – 1000
X2	Продолжительность	50 мин – 5 ч	50 – 300
X3	Жанр	Боевик; Приключение; Драма; Комедия; Криминальный; Мистика; Ужасы; Вестерн; Исторический; Биография; Анимация; Фантастика; Триллер; Мюзикл; Нуар	1 – 14
X4	Возрастное ограничение	0+; 6+; 12+; 16+; 18+; G; PG; PG-13; R; NC-17	1 – 5
X5	Франшиза	Да; Нет	0/1
X6	Сезон выхода	Зима; Весна; Лето; Осень	1 – 4
X7	Период высокой посещаемости	Да; Нет	0/1
X8	Рейтинг режиссера	0 – 10	0 – 10
X9	Наличие у режиссера престижных наград	Да; Нет	0/1
X10	Наличие у сценаристов престижных наград	Да; Нет	0/1
X11	Наличие у 3-х актеров на главных ролях престижных наград	Да; Нет	0/1
X12	Суммарное количество «оскаров» у съемочной группы	0 – 20	0 – 20



«Пермская научная школа искусственного интеллекта»

Прогноз кассовых сборов фильма

Бюджет фильма (доллары США):	<input type="text" value="Введите значение"/>
Возраст, от которого допускается просмотр фильма (согласно российской возрастной классификации информационной продукции):	<input type="text" value="Выберите значение"/>
Продолжительность фильма (минут):	<input type="text" value="Введите значение"/>
Наличие у режиссера номинаций/наград на престижные кинопремии («Оскар», «Золотой глобус», SAAG, Critics' Choice Awards):	<input type="text" value="Выберите значение"/>
Наличие у актеров номинаций/наград на на престижные кинопремии («Оскар», «Золотой глобус», SAAG, Critics' Choice Awards):	<input type="text" value="Выберите значение"/>
Основной жанр:	<input type="text" value="Выберите значение"/>
Наличие спецэффектов:	<input type="text" value="Выберите значение"/>
Является ли фильм продолжением какого-либо фильма (т.е. является ли фильм следующей серией франшизы – сиквелом):	<input type="text" value="Выберите значение"/>
Выход фильма в период высокой посещаемости кинотеатров (новогодние праздники, летние каникулы и т.д.):	<input type="text" value="Выберите значение"/>

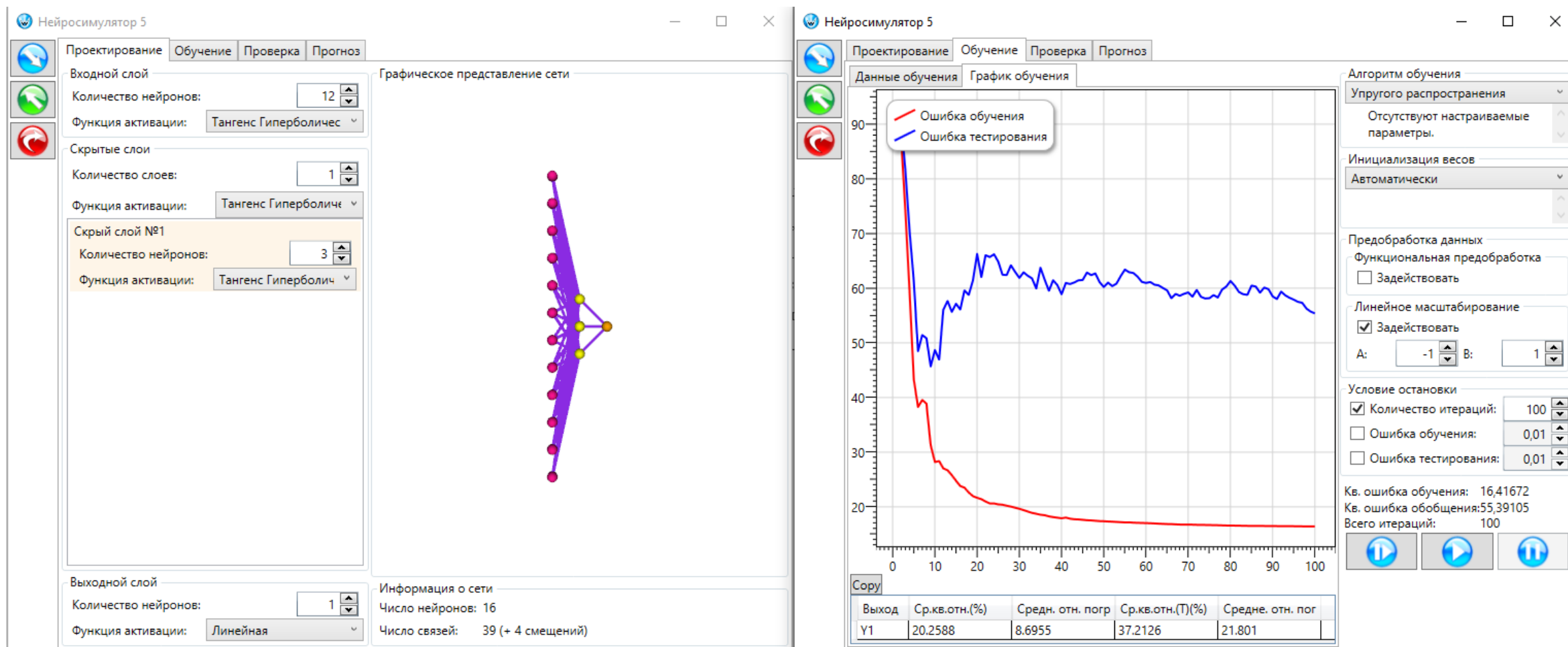


Рисунок 7 – Интерфейс программы «Нейросимулятор 5.0»



Модули:

1. Модуль сбора данных обучающего множества
2. Модуль очистки, систематизации и нормализации данных
3. Модуль нейронной сети
4. Пользовательский модуль программного продукта

Этапы:

1. Сбор данных для обучения.
2. Подготовка и нормализация данных.
3. Выбор топологии сети.
4. Экспериментальный подбор характеристик сети.
5. Экспериментальный подбор параметров обучения.
6. Тестовое обучение в программе «Нейросимулятор 5.0».
7. Обучение с использованием языка программирования.
8. Проверка адекватности обучения.
9. Корректировка параметров, окончательное обучение.
10. Вербализация нейронной сети с целью дальнейшего использования.



Входные данные

Для модуля очистки и модуля нейронной сети:

- Собранные данные модулем сбора данных в одном из форматов: «XSLX», «XLS», «CSV», «TXT».

Для клиентского приложения:

- Обученная нейронная сеть в формате «HDF5».
- Текстовые данные, введенные и/или выбранные пользователем.

Выходные данные

Для модуля очистки и систематизации:

- Файл в одном из форматов: «XSLX», «XLS», «CSV», «TXT».

Для модуля нейронной сети:

- Файл в формате «HDF5».

Для клиентского приложения:

- Результат в формате: успешно/не успешно.
- Числовые данные о примерных кассовых сборах.

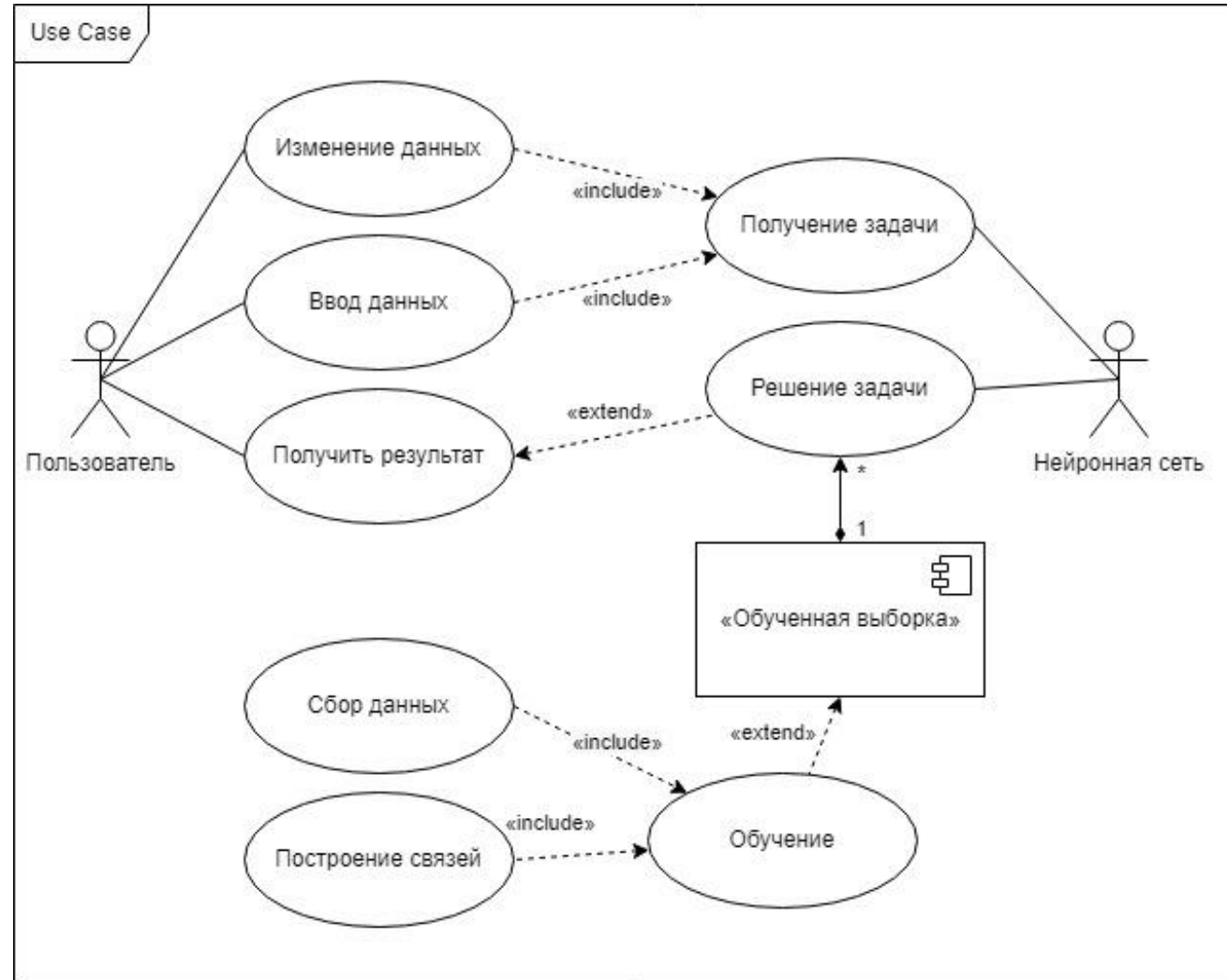
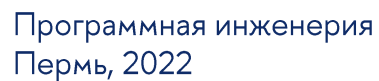


Рисунок 8 – Диаграмма вариантов использования



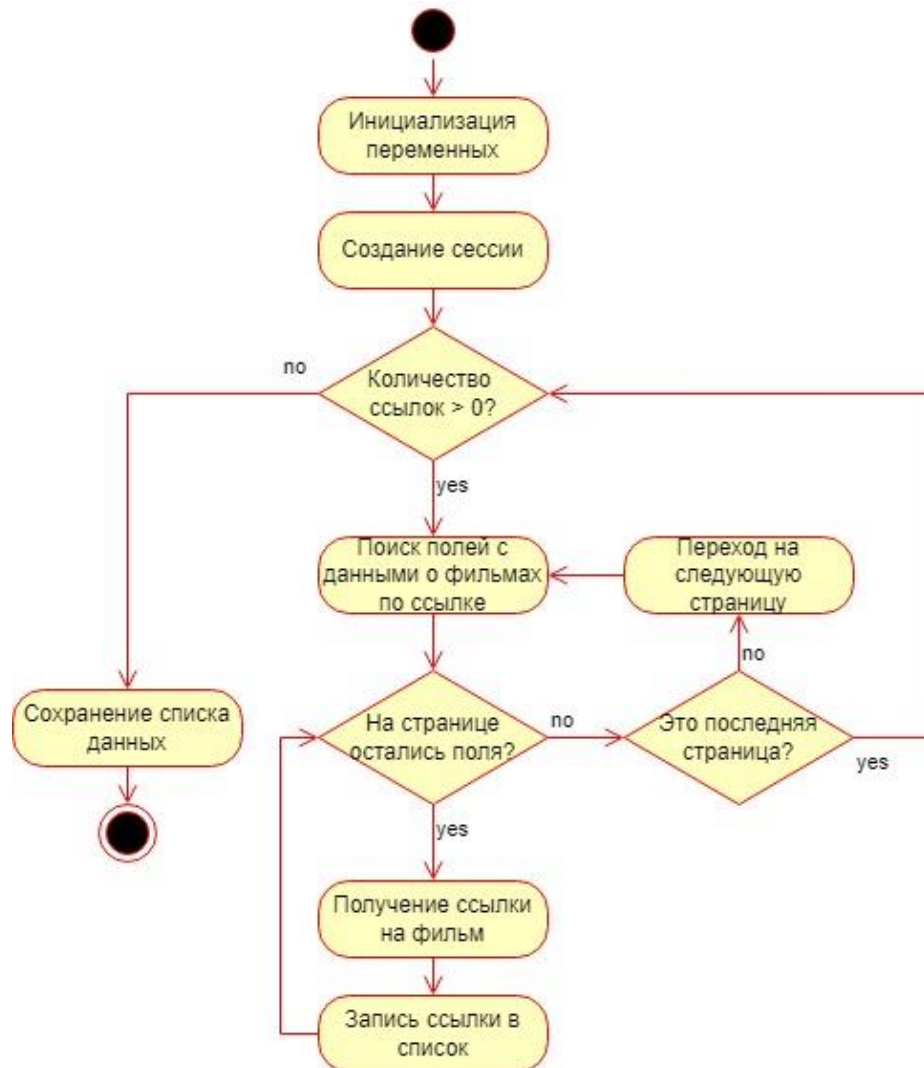


Рисунок 11 – Диаграмма активности алгоритма
первичного сбора данных

IMDb "Bottom 1000" (Sorted by IMDb Rating Descending)

1-50 of 1,000 titles. | [Next »](#)

View Mode: **Compact** | [Detailed](#)

Sort by: [Popularity](#) | [A-Z](#) | **[User Rating ▼](#)** | [Number of Votes](#) | [US Box Office](#) | [Runtime](#) | [Year](#) | [Release Date](#) | [Date of Your Rating](#) | [Your Rating](#)

	1. Сурьяванши (2021)	6,4	☆ Rate	+
	2. Бета-тест (I) (2016)	5,6	☆ Rate	+
	3. Мордекай (2015)	5,5	☆ Rate	+
	4. Обитатель звезд: Последний глаз (2016)	5,5	☆ Rate	+

Рисунок 12 – Рейтинг «BOTTOM – 1000»

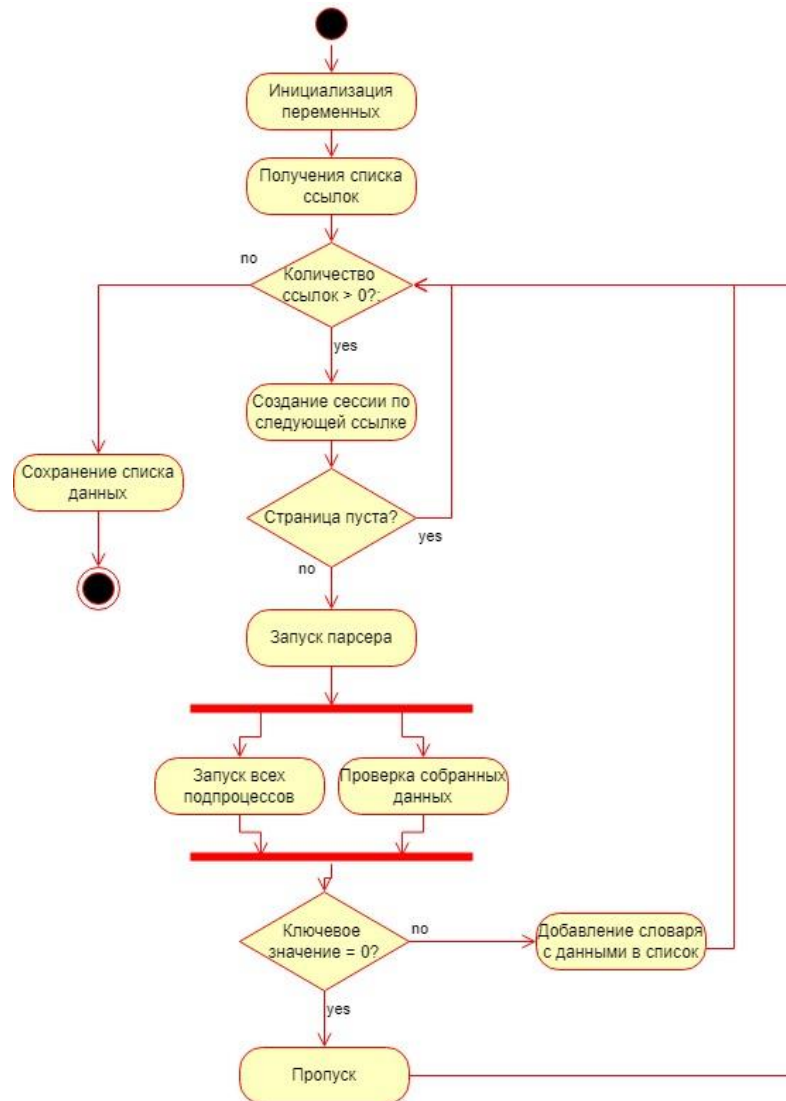


Рисунок 13 – Диаграмма активности алгоритма сбора параметров

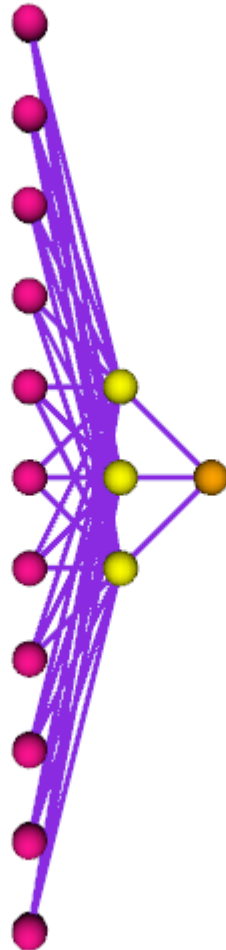


Рисунок 14 – Графическое представление нейронной
сети в программе «Нейросимулятор 5.0»

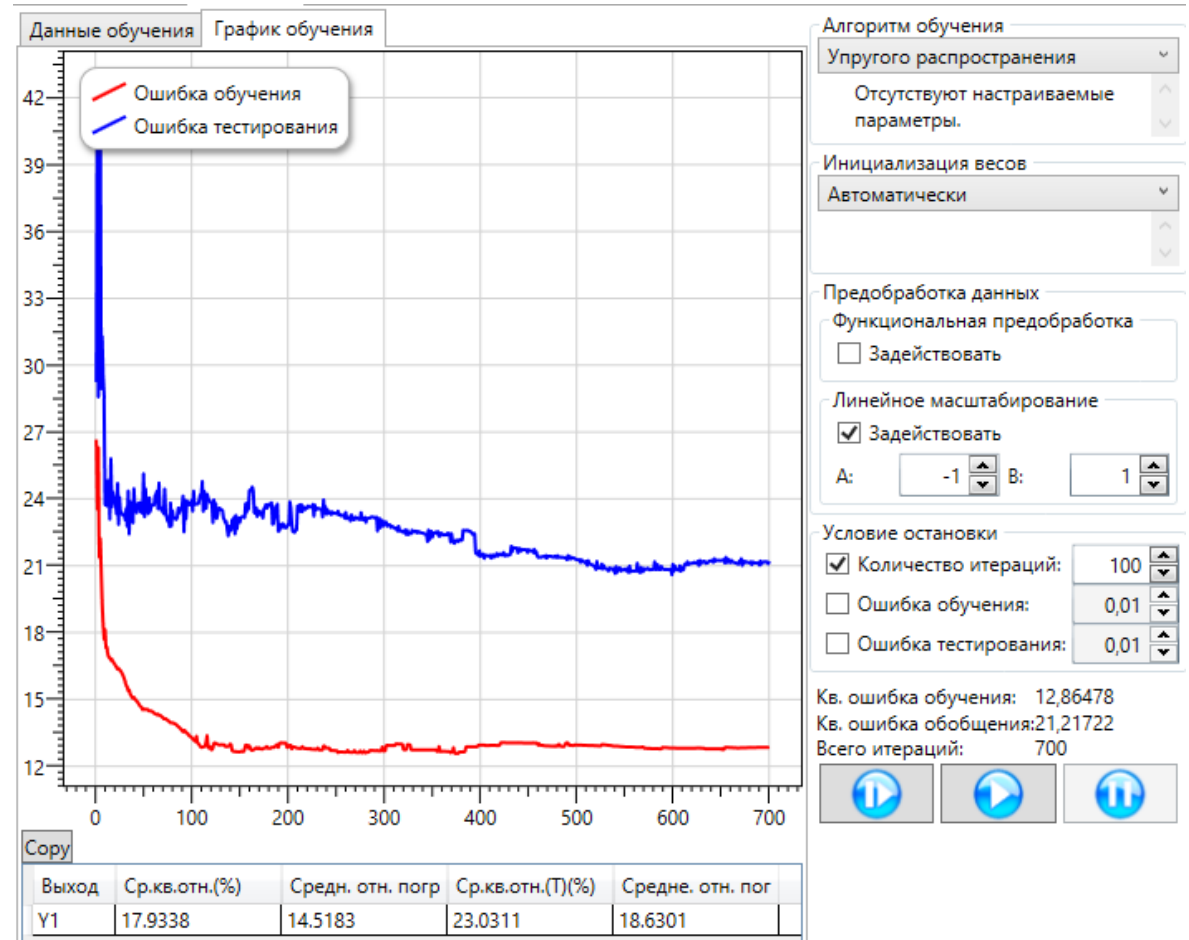


Рисунок 15 – График обучения нейронной сети в
программе «Нейросимулятор 5.0»

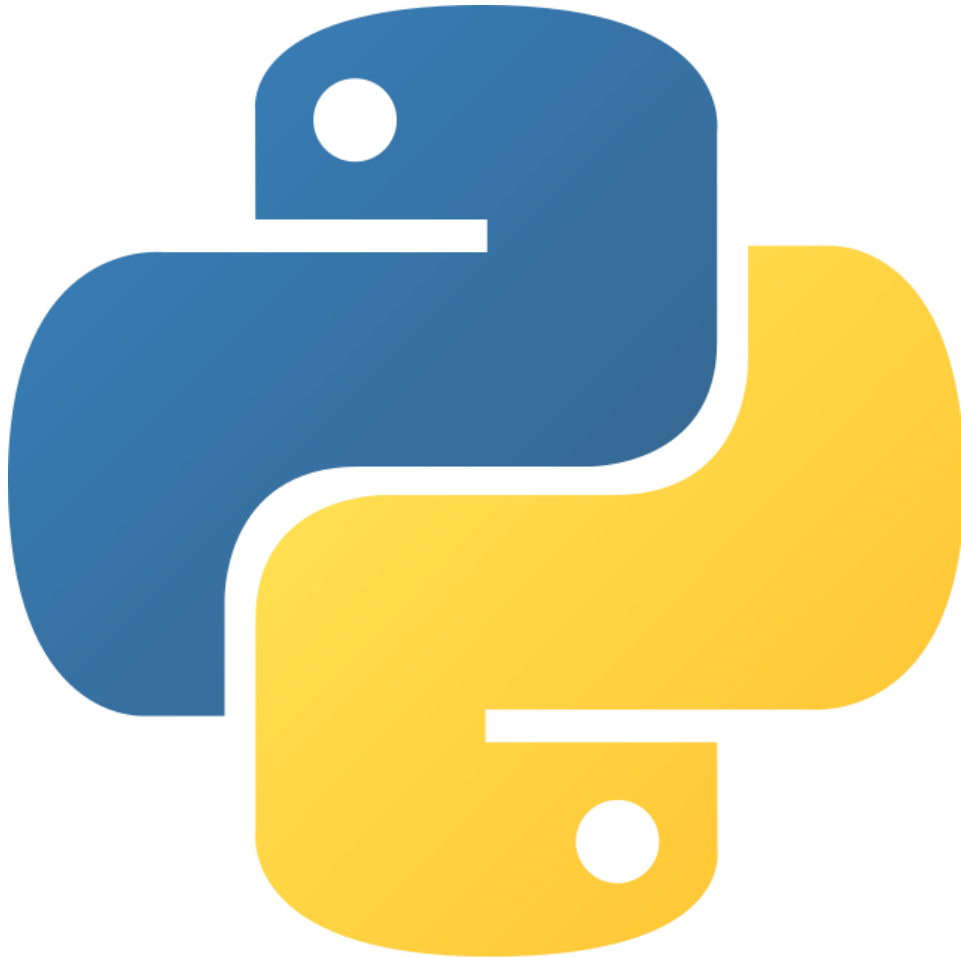


Рисунок 16 – Логотип языка Python



Рисунок 17 – Логотип приложения PyCharm



```
--- TEST PROCESSING ---
```

```
0/2000 Progress (0468569): [processing] https://www.imdb.com/title/tt0468569/
```

```
0/2000 Progress (0468569): [done] https://www.imdb.com/title/tt0468569/
```

```
ID:0468569 2008 Темный рыцарь
```

```
budget: 1800
```

```
duration: 152
```

```
genre: 1
```

```
age-limit: 3
```

```
franchise: 1
```

```
release-season: 2
```

```
holiday: 1
```

```
director-rating: 8
```

```
directors-awards: 1
```

```
writers-awards: 1
```

```
stars-awards: 1
```

```
oscars: 4
```

```
box-office: 10000
```

```
1800;152;1;3;1;2;1;8;1;1;1;4;10000
```

```
--- END IN 70.89 sec ---
```

Рисунок 18 – Процесс работы алгоритма сбора параметров

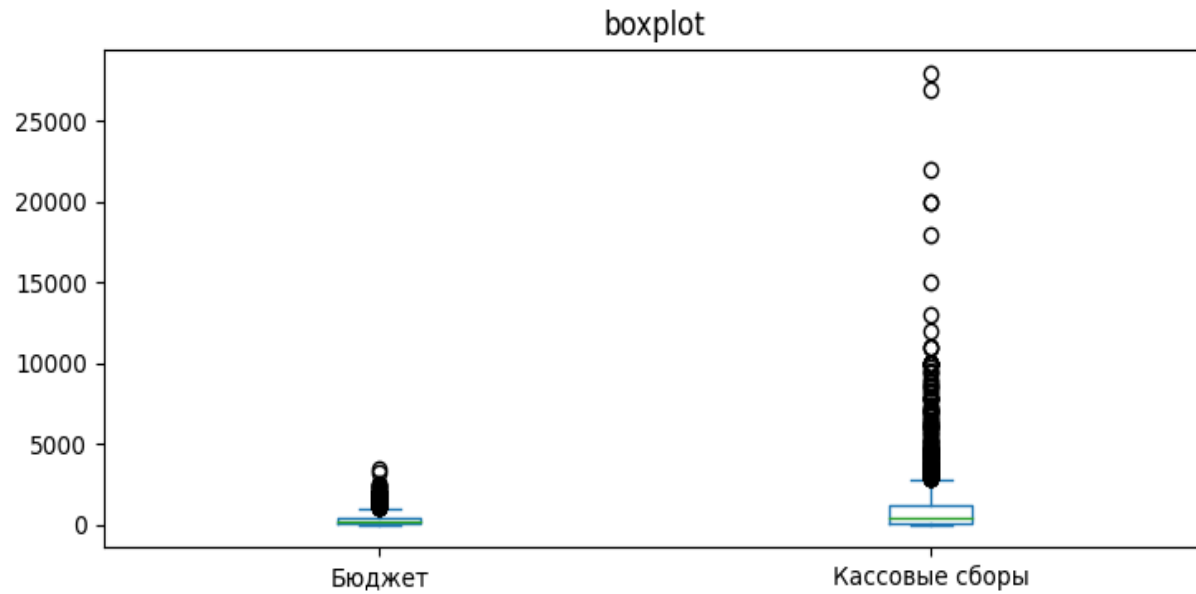


Рисунок 19 – Коробчатая диаграмма до удаления выбросов

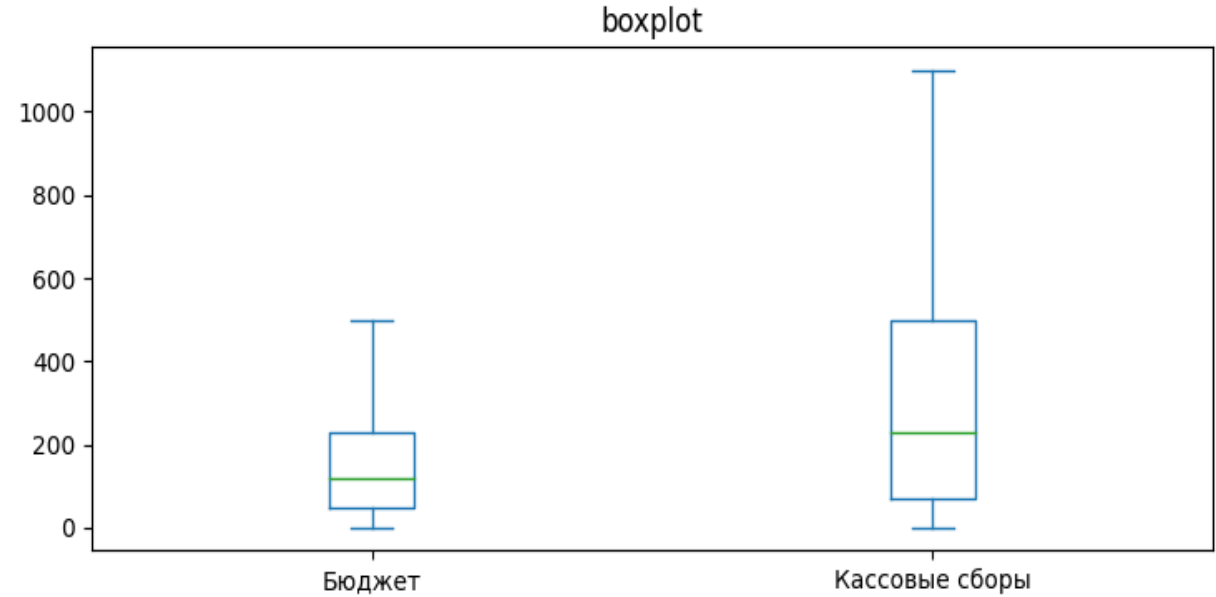


Рисунок 20– Коробчатая диаграмма после удаления выбросов

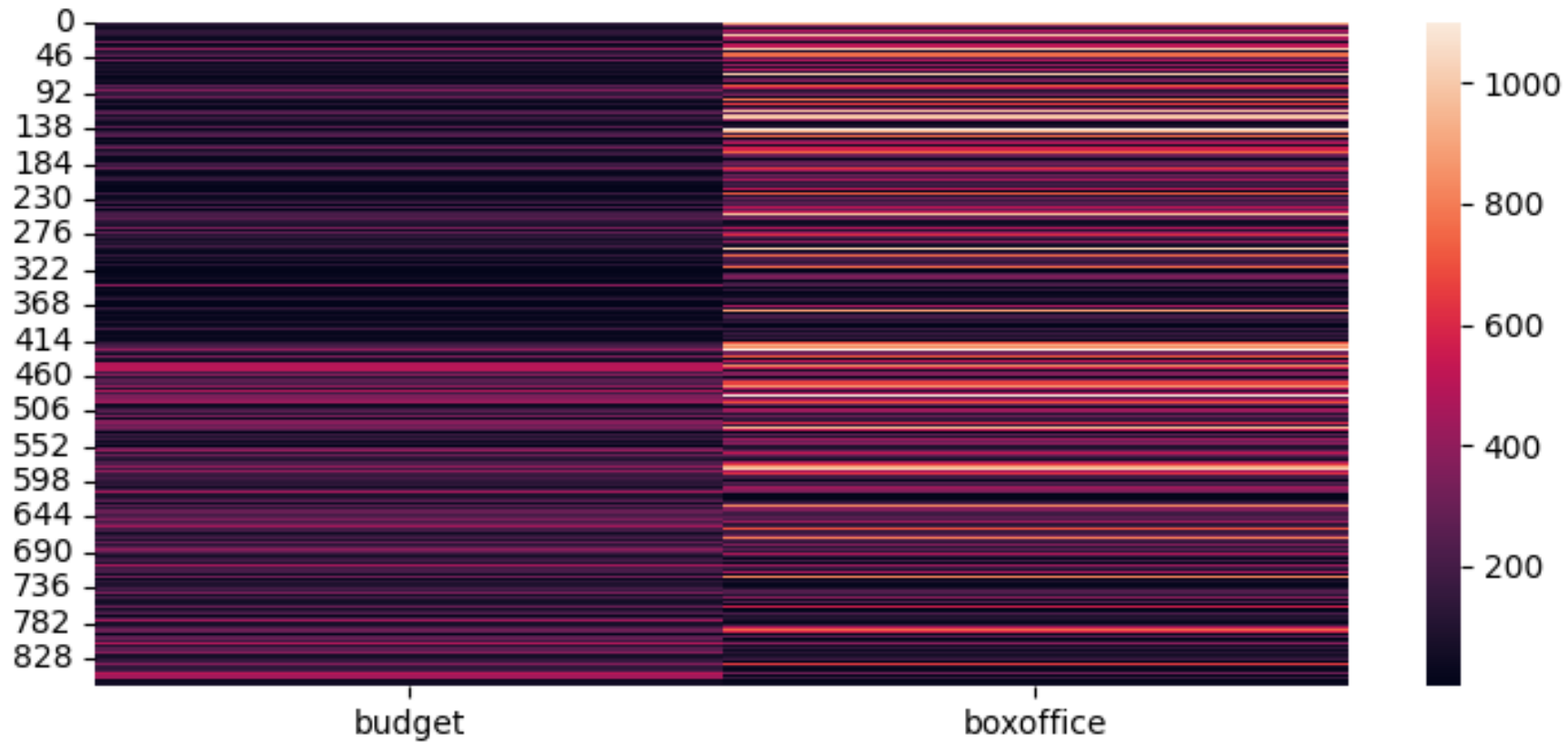


Рисунок 21 – Тепловая карта значений

error: 20.93%

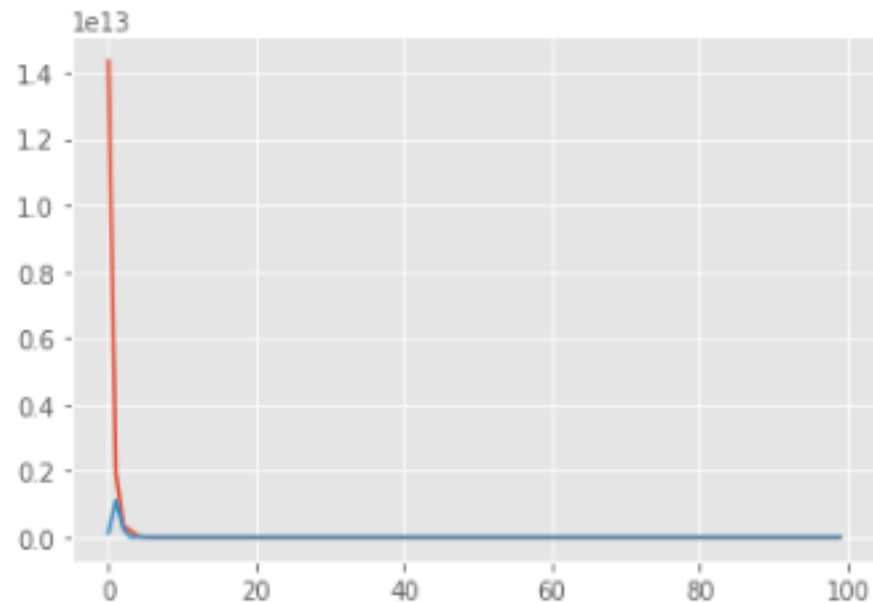


Рисунок 22 – Процесс обучения нейронной сети

$$E = \frac{\sqrt{\frac{\sum_{n=1}^N (d_n - y_n)^2}{N}}}{|\max(d_n) - \min(d_n)|} 100\%$$

Рисунок 23 – Формула расчета точности

d_n – фактические кассовые сборы

y_n – прогнозируемая величина

N – количество элементов выборки

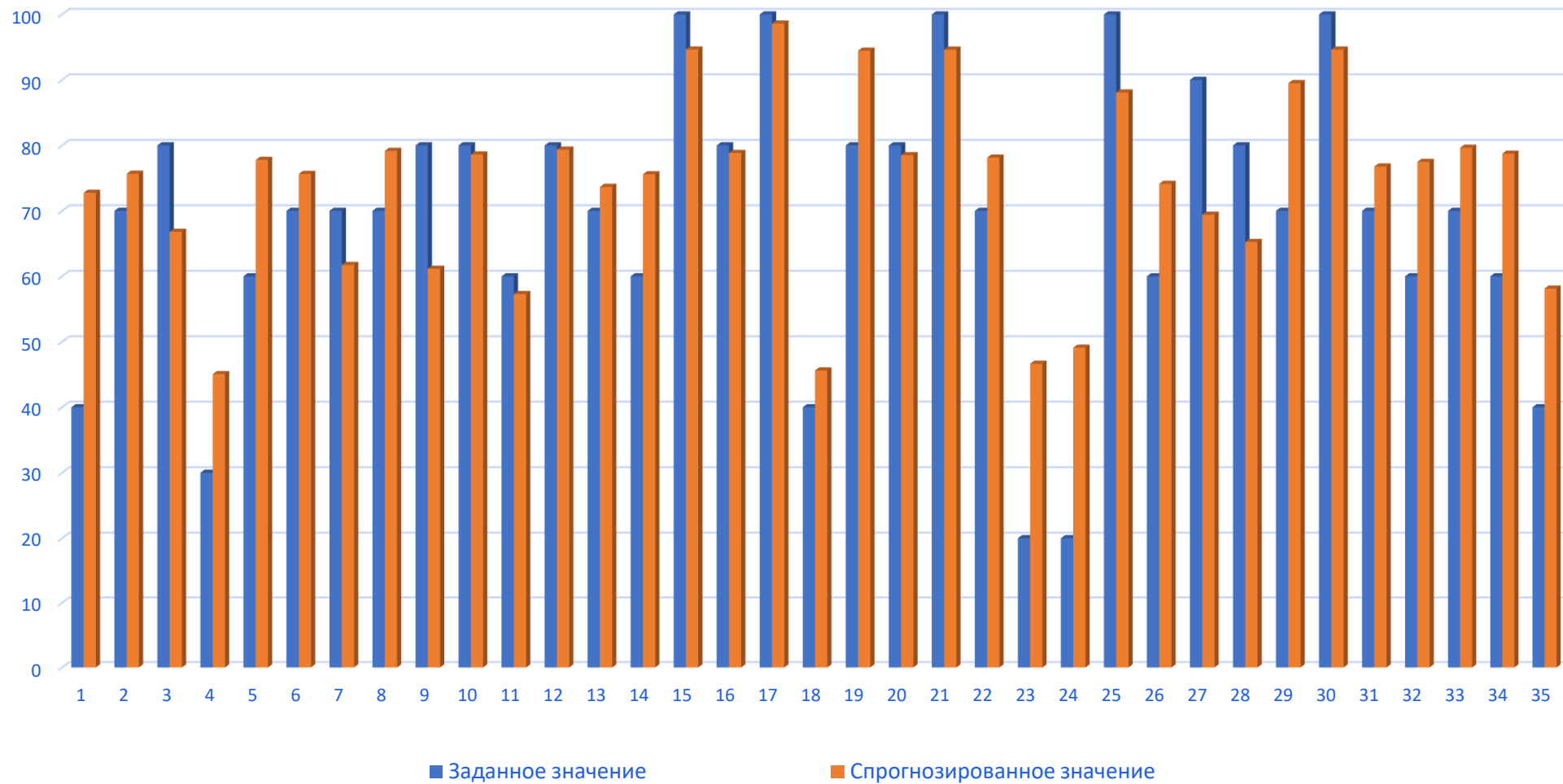


Рисунок 24 – Результат работы нейронной сети



Предсказание рентабельности кинокартины

Бюджет фильма (Доллары США)	<input type="text" value="100000"/>	Рассчитать
Продолжительность фильма (В минутах)	<input type="text" value="100"/>	
Основной жанр фильма	<input type="text" value="Боевик (Action)"/>	
Возрастное ограничение фильма	<input type="text" value="0+"/>	
Является ли фильм продолжением кинофраншизы (Является следующей частью уже вышедшего)	<input type="text" value="Нет"/>	
Планируемый сезон выхода	<input type="text" value="Зима"/>	
Планируется ли выход в период высокой посещаемости	<input type="text" value="Нет"/>	Прогнозируемые кассовые сборы: ~\$1000000
Рейтинг предыдущих работ режиссера	<input type="text" value="0"/>	
Имеются ли у режиссера престижные награды («Оскар», «Золотой глобус», SAAG, Critics' Choice Awards)	<input type="text" value="Нет"/>	
Имеются ли у сценариста престижные награды («Оскар», «Золотой глобус», SAAG, Critics' Choice Awards)	<input type="text" value="Нет"/>	
Имеются ли у актера престижные награды («Оскар», «Золотой глобус», SAAG, Critics' Choice Awards)	<input type="text" value="Нет"/>	
Суммарное количество оскаров у съемочной группы	<input type="text" value="10"/>	

Рисунок 25 – Интерфейс программы

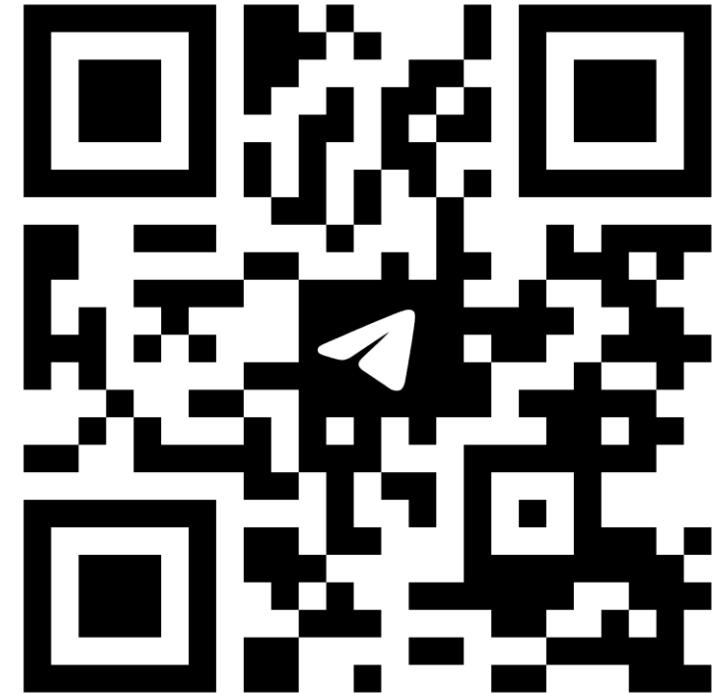


Рисунок 26 – Бот для прогнозирования рентабельности
кинобизнеса



1. Проведен анализ области машинного обучения и конкретизированы требования к системе
2. Нейронная сеть и модули системы спроектированы
3. Нейронная сеть и MVP система разработаны и протестированы

Подготовлена научная статья, представленная на «VIII всероссийскую научно-практическую конференцию с международным участием «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В РЕШЕНИИ АКТУАЛЬНЫХ СОЦИАЛЬНЫХ И ЭКОНОМИЧЕСКИХ ПРОБЛЕМ XXI ВЕКА»

СПАСИБО ЗА ВНИМАНИЕ

РАД ОТВЕТИТЬ НА ВАШИ ВОПРОСЫ



Автор: Чепокоев Е.С.

Телефон: +7 (951) 95-94-666

Электронная почта: escherokov@edu.hse.ru

