

УДК 004.021

## СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ОЦЕНКЕ КРЕДИТНЫХ РИСКОВ

Чуб Вадим Сергеевич<sup>1</sup>,  
e-mail: vadim-chub13@mail.ru,

<sup>1</sup>Донской государственный технический университет, г. Ростов-на-Дону, Россия

*В статье исследуется сравнительная эффективность автоматизированной оценки кредитных рисков современными методами машинного обучения (наивный байесовский классификатор, k-ближайших соседей, логистическая регрессия, «случайный лес» и глубокая нейронная сеть). Смоделирована база данных из легальных доступных открытых источников, произведены выборки, предварительная обработка данных. Выполнены обучение и тестирование моделей машинного обучения, их сравнение по метрикам качества. По результатам тестирования оптимизированы параметры модели машинного обучения полным перебором комбинаций статистических моделей (метода на основе правил, метода k-ближайших соседей, логистической регрессии, дискриминантного анализа, наивного байесовского классификатора, нейронных сетей и деревьев решений). Каждый из традиционных классификаторов (наивный байесовский классификатор, метод k-ближайших соседей, логистическая регрессия, «случайный лес», глубокая нейронная сеть) улучшался с применением механизмов стандартизации и снижения размерности методом главных компонент. По итогам сравнительного анализа моделей машинного обучения наилучший результат продемонстрировала модель нейронной сети с оптимизированными параметрами, обеспечившая наилучшие показатели по всем метрикам оценки.*

**Ключевые слова:** искусственный интеллект, машинное обучение, дропаут, краткосрочная память, матрица корреляции, кредитные риски

## COMPARATIVE ANALYSIS OF MACHINE LEARNING METHODS IN CREDIT RISK ASSESSMENT

Tshub V. S.<sup>1</sup>,  
e-mail: vadim-chub13@mail.ru,

<sup>1</sup>Don State Technical University, Rostov-on-Don, Russia

*The article describes the comparative effectiveness of automated credit risk assessment by modern machine learning methods (Naive Bayes classifier, k-nearest neighbors, logistic regression, “random forest” and deep neural network). A database was modeled from available open sources, samples were made, and data was preprocessed. Machine learning models have been trained and tested, and compared by quality metrics. Based on the test results, the parameters of the machine learning model were optimized by a complete search of combinations of statistical models (rule-based method, k-nearest neighbor method, logistic regression, discriminant analysis, naive Bayes classifier, neural networks and decision trees). Each of the traditional classifiers (naive Bayes classifier, k-nearest neighbor method, logistic regression, “random forest”, deep neural network) was improved with the use of standardization mechanisms and dimensionality reduction by the principal component method. Based on the results of a comparative analysis of machine learning models, the best result was demonstrated by a neural network model with optimized parameters, which provided the best indicators for all evaluation metrics.*

**Keywords:** artificial intelligence, machine learning, dropout, short-term memory, correlation matrix, credit risks

DOI 10.21777/2500-2112-2023-3-81-92

## Введение

Искусственный интеллект (ИИ) и особенно глубокое обучение – неотъемлемые компоненты многих технологий и сервисов современного цифрового общества. Среди причин – увеличение вычислительных мощностей, достижения в области анализа больших данных и возможности посредством алгоритмов глубокого обучения обнаруживать скрытые взаимосвязи в больших массивах данных. Всё более актуальным становится использование искусственных нейронных сетей для анализа больших объемов данных в финансовой сфере. Достижения финансовых технологий способствовали появлению новых способов ведения бизнеса. Кредитные организации предлагают всё более прозрачные условия и вместе с этим тщательно отбирают заявки на получение кредита. Повышается потребность в улучшении автоматизированных методов поддержки принятия решений по кредитным заявкам, предоставлении кредитов или отказе. Всё более распространено применение методов интеллектуального анализа данных, в частности, алгоритмов машинного обучения.

Для выбора алгоритма, предназначенного для оценки кредитных рисков, необходимо определить, какие из моделей дают наилучшие результаты различения «хороших» и «плохих» заемщиков. Выбор можно осуществить, пройдя все этапы интеллектуального анализа данных [1] кредитных заявок, сравнив преимущества и недостатки различных методов машинного обучения (логистическая регрессия, дерево решений, «случайный лес», наивный байесовский классификатор, нейронные сети) для решения поставленной задачи.

Задачей данной статьи является исследование возможности применения методов машинного, в частности, глубокого обучения для автоматизации оценки кредитного риска и повышения точности оценки кредитных рисков.

Метод исследования – сравнительный анализ экспериментальных исследований различных моделей машинного обучения на имеющихся данных для автоматизированной генерации на основании ранее неизученных данных вывода, насколько податель заявки надёжен для выдачи кредита. Такой подход был успешно апробирован в ряде исследований [2; 3]. В статье проводится сравнение использования в моделях машинного обучения традиционных классификаторов: наивный байесовский классификатор, метод k-ближайших соседей, логистическая регрессия, «случайный лес» и глубокая нейронная сеть. Выбор этих методов обусловлен их популярностью в научном сообществе применительно к задачам автоматизированной оценки кредитных рисков [4].

Сравнительное исследование семи методов интеллектуального анализа данных в кредитном скоринге (статистической модели – метода на основе правил, метода k-ближайших соседей, логистической регрессии, дискриминантного анализа, наивного байесовского классификатора, нейронных сетей и деревьев решений) показало лучшие результаты для нейронных сетей [2]. Анализ рисков клиентов кредитных карт проводился при помощи методов опорных векторов, k-ближайших соседей, дерева решений и «случайного леса» с нейронной сетью прямого распространения [3]. Для повышения точности результатов автором предлагалось добавить в нейронные сети два важных фактора – дропаут (исключение случайной доли данных в процессе обучения сети) и долгую краткосрочную память – чтобы найти их влияние на повышение точности, а также решить проблему переобучения [2]. Точность (доля верных откликов модели) нейронных сетей в проведённом исследовании составила 0.8246, превзойдя другие методы. Отмечено, что использование дропаута повышает производительность нейронных сетей прямого распространения, в особенности при увеличении численности нейронов в слое. Оговаривается, что использование дропаута в долгой краткосрочной памяти (LSTM) практически не увеличивает точность.

Новизна исследования, представленного в настоящей статье, заключается в предложении комбинированного способа различных методов машинного обучения для получения, анализа и обработки данных в условиях дефицита легальных публичных кредитных заявок, обеспечивающего повышение точности оценки кредитных рисков.

Так, ряд исследователей использовали собственные массивы данных конфиденциального характера [2; 3; 5], и публикация могла усилить уязвимость клиентов компании.

Результаты исследования и их обсуждение

В экспериментальной части настоящего исследования применён метод моделирования набора данных с атрибутами части клиентских кредитных заявок. Создание искусственно смоделированной клиентской базы данных проводилось посредством библиотеки Russian Names<sup>1</sup>, позволяющей генерировать реалистичные русские фамилии, имена и отчества (ФИО). При проведении эксперимента использовалась платформа Google Colab и популярные библиотеки машинного обучения, анализа и визуализации данных (NumPy, Matplotlib, Seaborn, Scikit-Learn, Pandas и Keras)<sup>2</sup>.

Из-за необходимых для моделирования кредитного скоринга массивов реальных кредитных данных использовался общедоступный набор Kaggle<sup>3</sup>. В нём содержатся столбцы, имитирующие данные кредитной финансовой организации. В таблице 1 приведено объяснение значений (полей), набор данных содержит 12 атрибутов (признаков) и 32 581 экземпляр (строку) – как категориальные, так и числовые. Целевой переменной в этом наборе является loan\_status, указана информация о клиентах, взявших кредит в неуказанном финансовом учреждении. Задачей анализа данных является поиск статистических взаимосвязей, позволяющих выявлять кредитные риски, разрабатывать алгоритмы прогнозирования кредитного дефолта.

Таблица 1 – Описание признаков в наборе данных<sup>4</sup>

Название признака	Значение признака	Описание признака
person_age	целое	Возраст, лет
person_income	целое	Годовой доход (долларов США)
person_home_ownership	категориальное	Тип собственности
person_emp_length	вещественное	Занятость клиента (месяцев)
loan_intent	категориальное	Цель кредита
loan_grade	категориальное	Оценка кредита
loan_amnt	целое	Сумма кредита (долларов США)
loan_int_rate	вещественное	Ставка по кредиту (% годовых)
loan_status	бинарное	Статус (дефолт по кредиту или не дефолт)
loan_percent_income	вещественное	Проценты по кредиту
cb_person_default_on_file	категориальное	Исторический дефолт (был ли невозврат кредита ранее)
cb_person_cred_hist_length	целое	Продолжительность кредитной истории (месяцев)

Учитывая контекст сбора данных, невозможно учитывать внешние макроэкономические события, которые могли бы существенно повлиять на результаты анализа. Например, в разных странах могут быть разные потребности граждан в кредитах, разные сценарии развития и завершения экономических циклов. Считается, что набор данных не учитывает кредитный рейтинг. То есть до получения данных систематическая проверка кредитоспособности клиентов не проводилась.

Переменная person\_home\_ownership имеет четыре значения: ипотека (Mortgage), собственность (Own), аренда (Rent) и другое (Others).

Переменная loan\_intent имеет шесть различных значений: консолидация долга (Debt consolidation), образование (Education), улучшение жилищных условий (Home improvement), медицина (Medical), личное (Personal), предприятие (Venture).

Переменная credit\_grade представляет собой оценку, которая может иметь одно из семи различных значений: A, B, C, D, E, F, G. Конкретное значение учитывает комбинацию нескольких показателей кредитного риска из кредитного отчета и кредитной заявки. Эти факторы могут включать уровень поддержки поручителей, историю погашения, денежный поток, прогнозируемые ежегодные расходы и т.д.

<sup>1</sup> Russian Names | PIP. – URL: <https://pypi.org/project/russian-names/> (дата обращения: 01.06.2023). – Текст: электронный.

<sup>2</sup> Machine Learning libraries (NumPy, SciPy, matplotlib, scikit-learn, pandas). – URL: <https://www.dotnetlovers.com/Article/217/machine-learning-libraries-numpy-scipy-matplotlib-scikit-learn-pandas> (дата обращения: 31.05.2023). – Текст: электронный.

<sup>3</sup> Credit Risk Dataset | Kaggle. – URL: <https://www.kaggle.com/datasets/laotse/credit-risk-dataset> (дата обращения: 02.06.2023). – Текст: электронный.

<sup>4</sup> Разработано автором по данным: Credit Risk Dataset | Kaggle (<https://www.kaggle.com/datasets/laotse/credit-risk-dataset>).

Переменная `cb_person_default_on_file` имеет только два значения: Y, если у клиента уже был дефолт по кредиту, N – если не было.

Для автоматизации оценки кредитного риска были выбраны несколько атрибутов из описанного выше набора данных:

- `person`, как «ФИО» (сгенерированные данные клиента);
- `person_age`, (данные клиента);
- `loan_int_rate`;
- `cb_person_cred_hist_length`, как «Срок» (количество месяцев);
- `person_income`, как «Доход» (в условных денежных единицах в год);
- `loan_amnt`;
- `cb_person_default_on_file`, как «Невозврат» (0 – не было дефолтов, невыплаченных платежей по предыдущим кредитам, 1 – были невыплаченные платежи);
- `loan_status`, как «Риск» (целевая переменная: 0 – низкий риск, 1 – высокий риск).

Таким образом, модель машинного обучения, предназначенная для прогнозирования кредитного риска, должна определить: отказать клиенту в кредите по представленным данным или одобрить заявку.

На рисунке 1 приведена круговая диаграмма распределения значений целевой переменной.



Рисунок 1 – Круговая диаграмма распределения значений целевой переменной<sup>5</sup>

Из рисунка 1 следует, что классы в наборе данных не сбалансированы – метку «0», соответствующую отсутствию кредитного риска или низкому риску, имеет 23001 строка в наборе данных (около 78 % от всех записей в наборе данных), а метку «1», соответствующую высокому риску невозврата кредита, только 6464 строки (около 22 % от всех записей).

Ранее невыплаченные кредиты обнаруживаются примерно у 18 % клиентов, как показано на круговой диаграмме, приведенной на рисунке 2.

Результат вывода описательных характеристик по каждому атрибуту в наборе данных приведен на рисунке 3.

Здесь выполняется подсчет количества значений по каждому атрибуту, вывод среднего значения, вычисление стандартного отклонения, минимума и максимума, а также так называемых процентилей (25, 50 и 75). Значения большинства переменных в наборе данных распределены почти равномерно с небольшим количеством выбросов, о чем свидетельствуют невысокие показатели 25-го либо 75-го процентиля.

Тепловые карты или матрицы корреляции позволяют отследить зависимости между числовыми признаками в наборе данных – как между парами атрибутов, так и в отношении каждого атрибута к целевой переменной. Коэффициент корреляции может быть вещественным числом от  $-1$  до  $+1$ . Значения, близкие к  $+1$ , означают, что между двумя парами переменных обнаруживается сильная положительная корреляция. В противном случае, если значения близки  $-1$ , переменные имеют сильную отрицательную корреляцию. Рисунок 4 представляет матрицу корреляции для набора данных.

<sup>5</sup> Составлено автором.

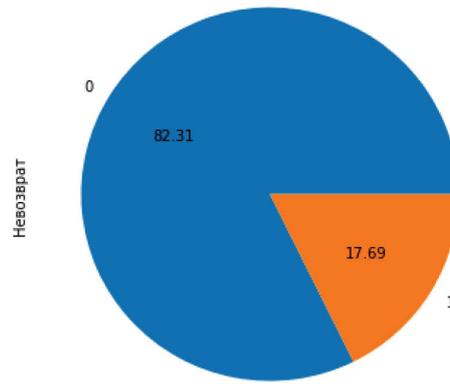


Рисунок 2 – Круговая диаграмма распределения значений переменной<sup>6</sup>

	count	mean	std	min	25%	50%	75%	max
<b>Возраст</b>	29465.0	27.714712	6.300193	20.00	23.0	26.00	30.00	144.00
<b>Ставка</b>	29465.0	11.011695	3.240459	5.42	7.9	10.99	13.47	23.22
<b>Срок</b>	29465.0	5.788257	4.031987	2.00	3.0	4.00	8.00	30.00
<b>Доход</b>	29465.0	66020.470490	61901.422932	4000.00	38500.0	55000.00	79100.00	6000000.00
<b>Сумма</b>	29465.0	9584.744612	6316.272282	500.00	5000.0	8000.00	12250.00	35000.00
<b>Невозврат</b>	29465.0	0.176922	0.381609	0.00	0.0	0.00	0.00	1.00
<b>Риск</b>	29465.0	0.219379	0.413833	0.00	0.0	0.00	0.00	1.00

Рисунок 3 – Описательные характеристики в наборе данных<sup>7</sup>

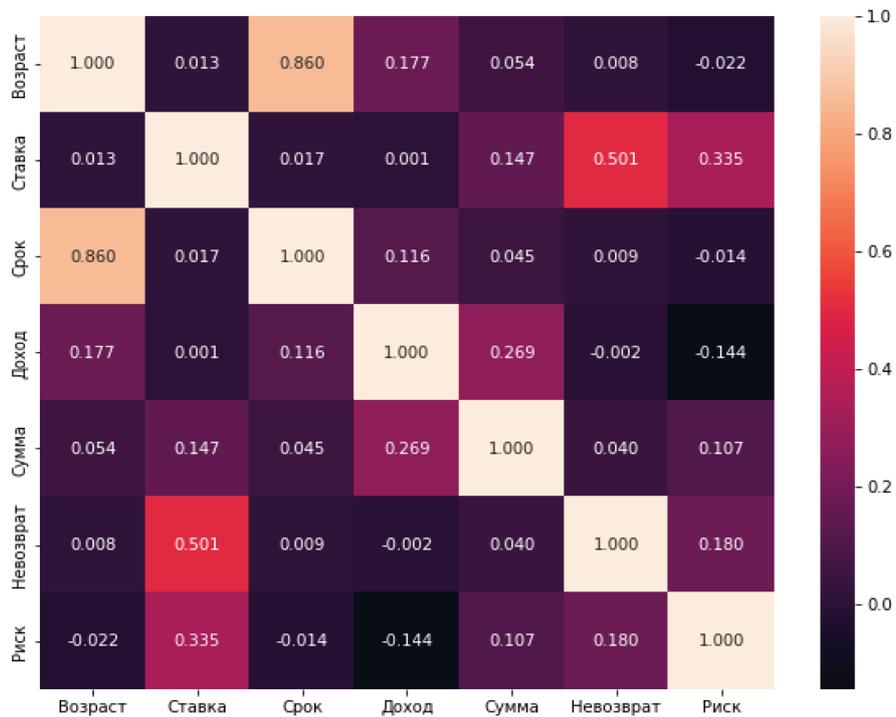


Рисунок 4 – Матрица корреляции для набора данных<sup>8</sup>

<sup>6</sup> Составлено автором.

<sup>7</sup> Составлено автором.

<sup>8</sup> Составлено автором.

Можно отметить, что на целевую переменную сильнее всего влияет атрибут «Ставка». Чем выше годовая ставка кредита, тем выше риск его невозврата.

Попарно сильно коррелируют следующие атрибуты:

- «Возраст» и «Ставка» (чем выше возраст клиента, тем выше ставка);
- «Ставка» и «Невозврат» (клиенты с высокой ставкой кредита чаще не возвращали кредиты ранее);
- «Доход» и «Сумма» (чем выше доход клиента, тем больше сумма кредита).

Также можно отметить, что несколько атрибутов имеют хорошую корреляцию в отношении атрибута «Риск» – целевой переменной кредитного риска. Риск невозврата кредита предсказуемо растёт с увеличением таких параметров, как «Ставка» и «Сумма». Кроме того, «Невозврат» повышает риск неплатежей по кредиту. Атрибуты «Срок» и «Возраст» незначительно влияют на зависимую переменную. Атрибут «Доход» коррелирует отрицательно, со снижением дохода клиента риск невозврата кредита увеличивается.

Следующий этап – выявление аномальных значений (выбросов) в наборе данных. Для поиска выбросов в наборе данных используется так называемый межквартильный размах (Interquartile Range, IQR). Им измеряется разброс средних 50 % значений в наборе данных, для чего вычисляется разница между 25-м перцентилем (Q1) и 75-м перцентилем (Q3). Любое значение, выходящее за пределы диапазона от  $-1,5 \times IQR$  до  $+1,5 \times IQR$ , считается выбросом. Для визуализации IQR были построены блочные диаграммы (рисунок 5).

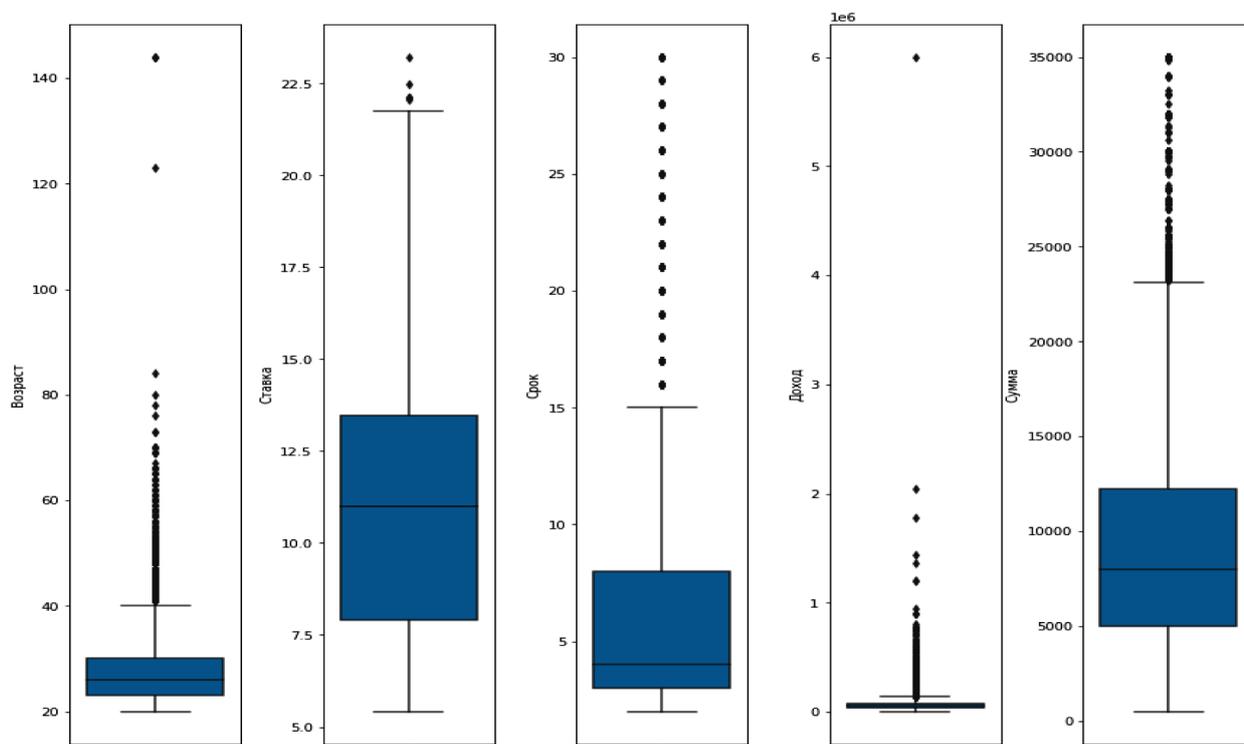


Рисунок 5 – Диаграмма 1. Блочные диаграммы в наборе данных<sup>9</sup>

Блочная диаграмма доходов клиентов показывает наличие выбросов в данных. Кроме того, она показывает, что средний доход клиентов без риска невозврата кредита выше, чем средний доход клиентов с риском невозврата. У клиентов с высоким риском сумма кредитов выше, чем у клиентов с низким риском. Самые большие различия между двумя классами клиентов обусловлены их средней процентной ставкой. Средняя процентная ставка по кредиту составляет примерно 10,43 % для клиентов без риска по кредиту и 13,06 % для клиентов с высоким кредитным риском.

<sup>9</sup> Составлено автором.

Как следует из диаграммы 2 (рисунок 6), выбросы обнаруживаются во всех рассмотренных столбцах. Было найдено 10 строк в наборе данных, у которых число выбросов больше трех: 31 810, 32 291, 31 972, 31 818, 32 305, 32 562, 32 312, 32 539, 32 124 и 32 541. В наборе данных после удаления выбросов осталось 29 455 записи.

Данные по-прежнему распределены по классам неравномерно (рисунок 6). В итоговом наборе данных 22 994 строки принадлежат классу 0 (низкий риск невозврата кредита), а 6461 строк – классу 1 (высокий риск).

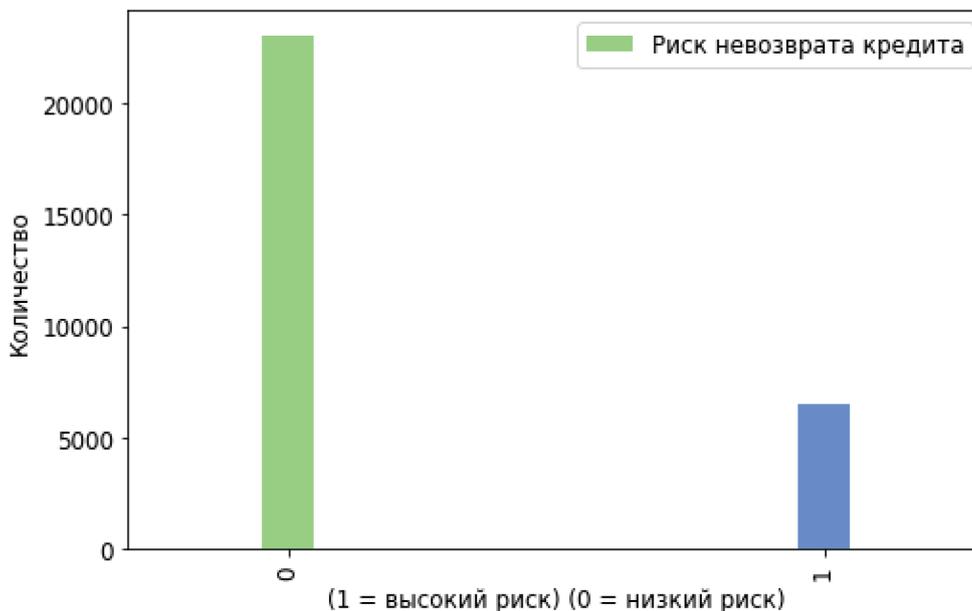


Рисунок 6 – Диаграмма 2. Распределение целевой переменной в наборе данных<sup>10</sup>

При оценке модели можно использовать метрику ассигасы, однако необходимо учитывать, что разные типы ошибок могут иметь разное значение. Поэтому приходится задействовать другие две метрики – точность (precision) и полноту (recall). Для устранения дисбаланса классов используется андерсэмплинг (большой класс заменяется подвыборкой, мощностью равной малому классу). В сбалансированном финальном наборе данных 12 922 строки, по 6461 для каждого класса [6–8].

Перед применением моделей машинного обучения набор данных был в соотношении 80/20 разбит на выборки обучающую (10 337 строк) и тестовую (2585 строк). Переменная «ФИО» не влияет на целевую переменную, поэтому её значения были удалены из набора данных, целевой переменной является «Риск».

Диаграммы 3–6 (рисунки 7–10) и схема 1 (рисунок 11) иллюстрируют графики сравнения полученных результатов для всех рассмотренных алгоритмов. Как следует из анализа результатов моделирования, наилучший результат показали модели на основе нейронной сети прямого распространения (глубокие перцептроны с несколькими скрытыми слоями), продемонстрировав самые высокие показатели по всем метрикам. Архитектура модели глубокой нейронной сети, построенной с использованием оптимальных гиперпараметров, приведена на схеме 1.

<sup>10</sup> Составлено автором.

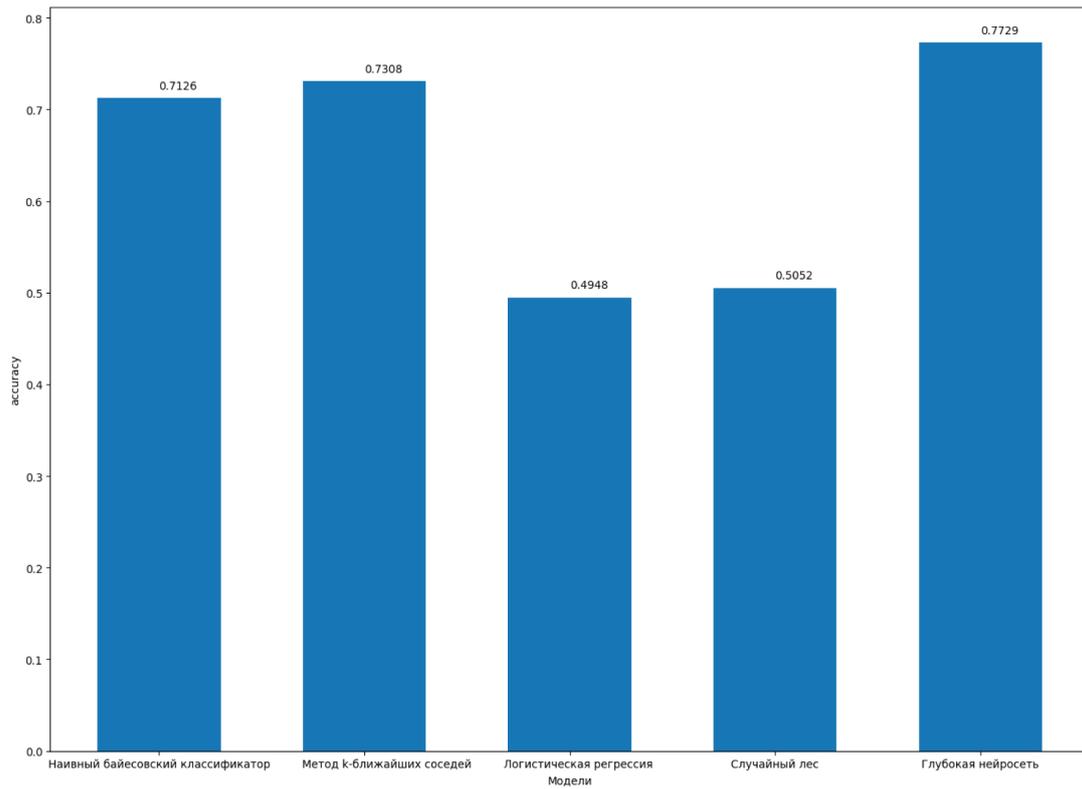


Рисунок 7 – Диаграмма 3. Сравнение алгоритмов машинного обучения по доле верных откликов<sup>11</sup>

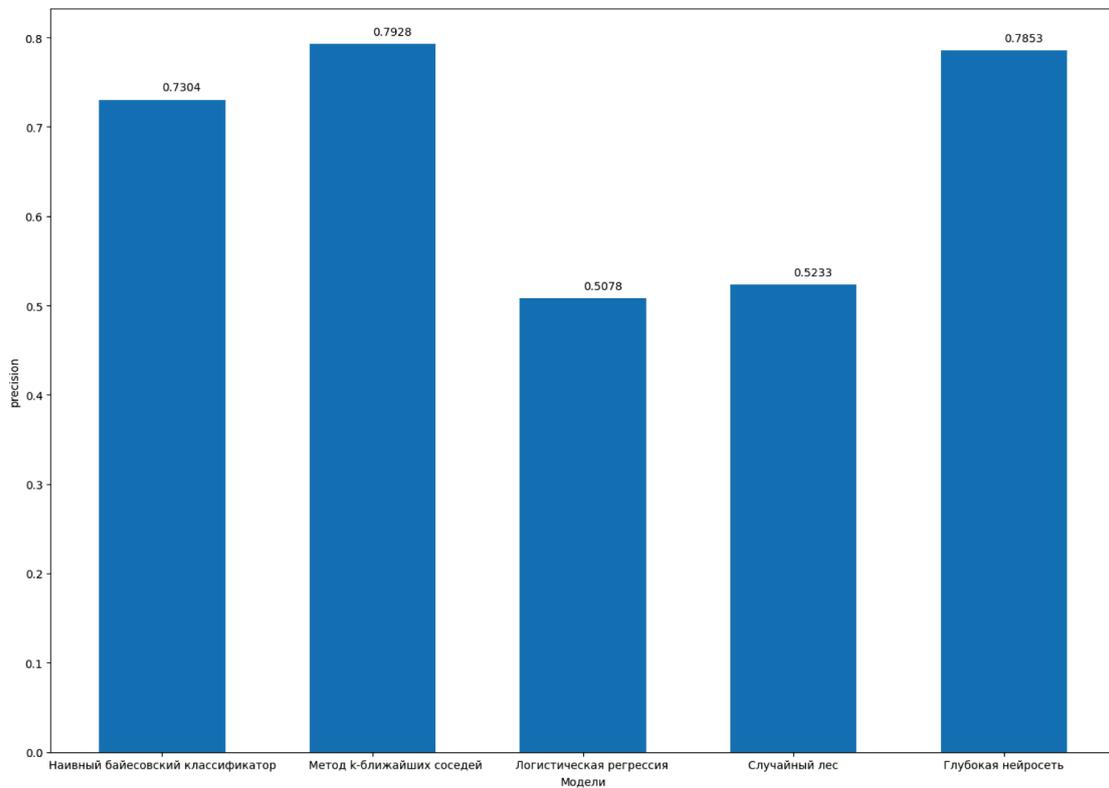


Рисунок 8 – Диаграмма 4. Сравнение алгоритмов машинного обучения по точности<sup>12</sup>

<sup>11</sup> Составлено автором.

<sup>12</sup> Составлено автором.

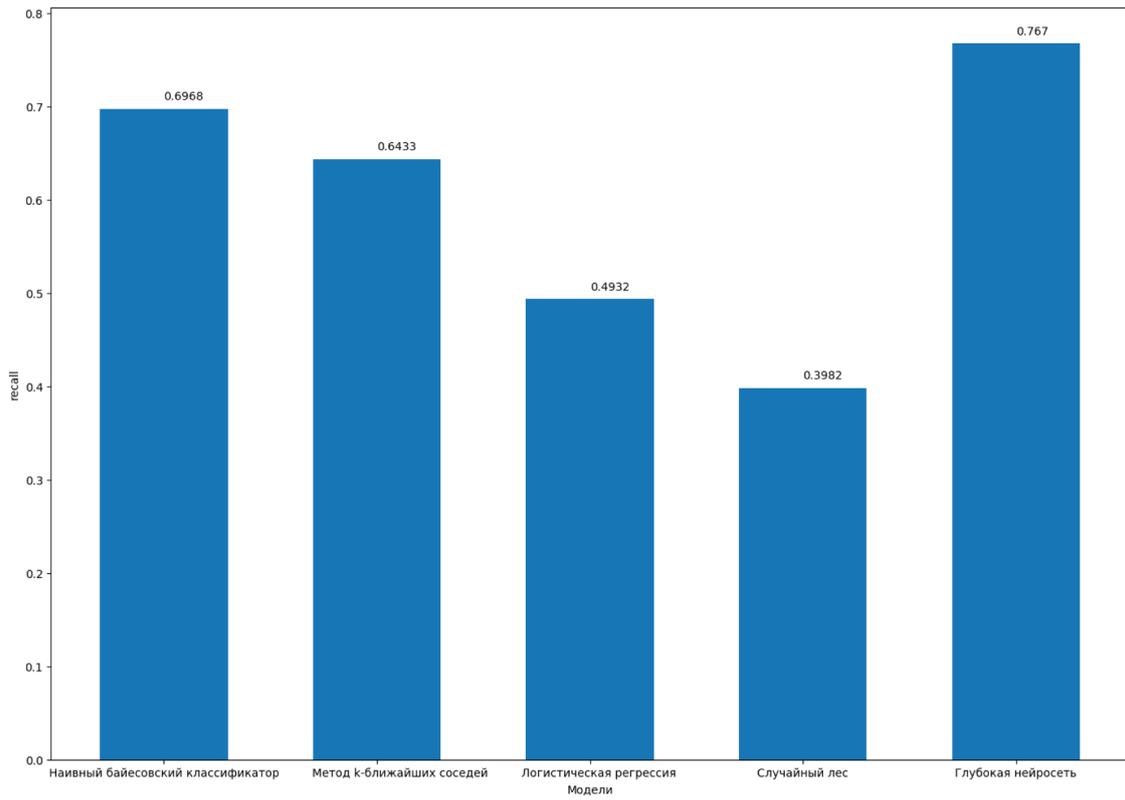


Рисунок 9 – Диаграмма 5. Сравнение алгоритмов машинного обучения по полноте<sup>13</sup>

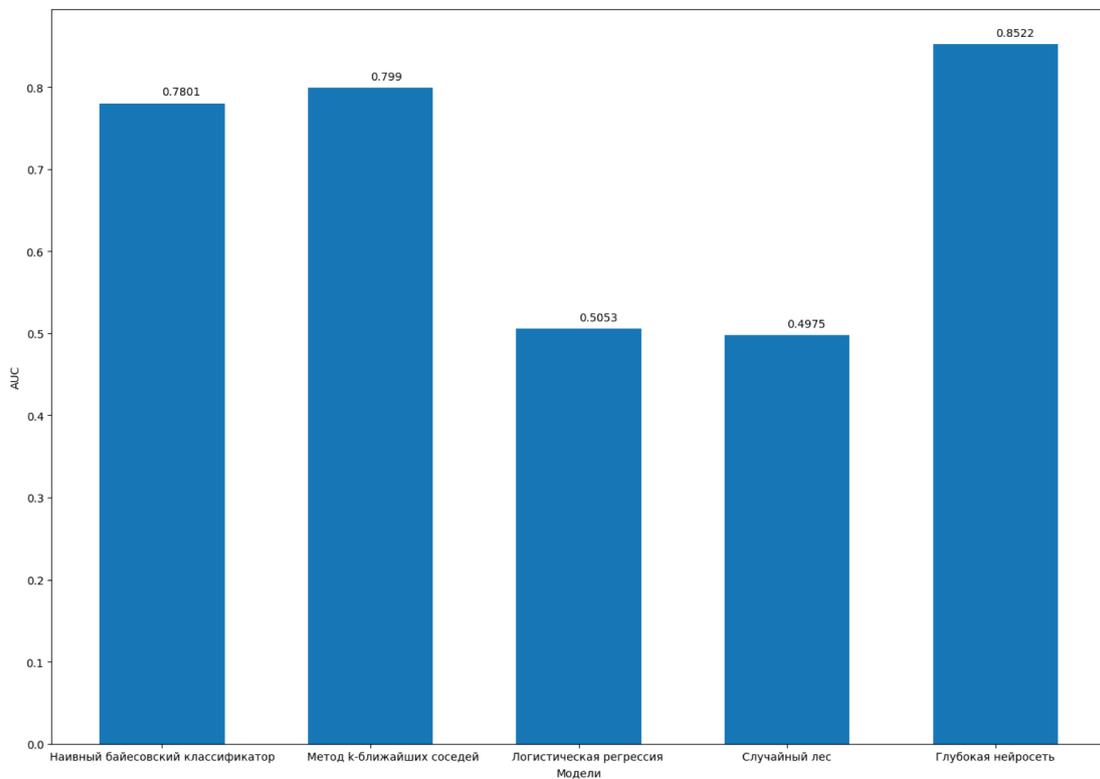


Рисунок 10 – Диаграмма 6. Сравнение алгоритмов машинного обучения по величине AUC<sup>14</sup>

<sup>13</sup> Составлено автором.

<sup>14</sup> Составлено автором.

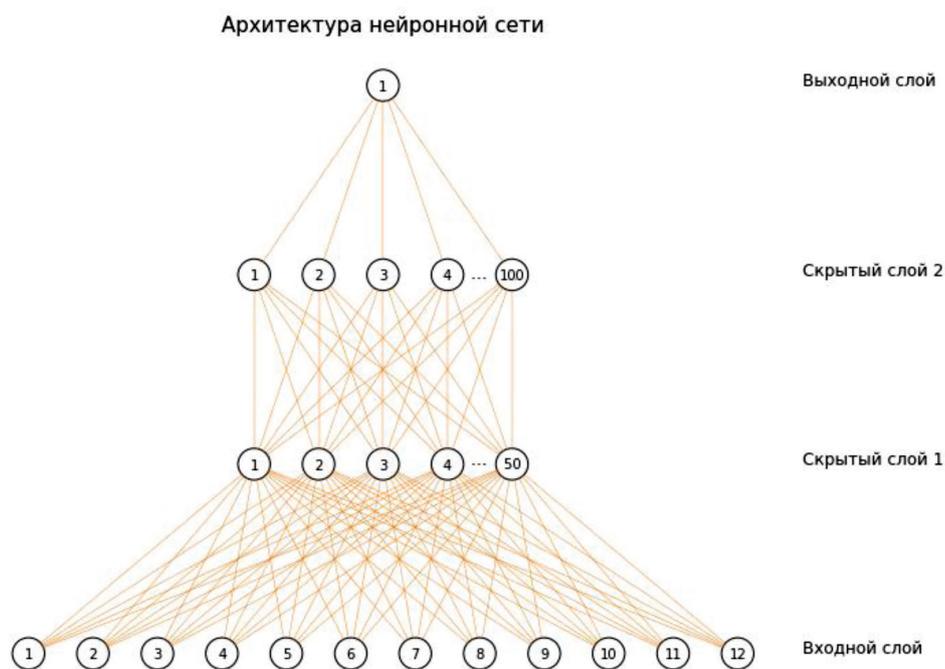


Рисунок 11 – Схема 1. Архитектура модели глубокого обучения для решения задачи с оптимальными гиперпараметрами<sup>15</sup>

Лучшая из моделей нейронной сети может быть сохранена в отдельный файл для дальнейшей работы в пользовательском приложении, которое позволяет выполнить автоматизацию оценки кредитных рисков.

### Заключение

В представляемом в статье исследовании решена задача автоматизации оценки кредитных рисков на смоделированном наборе данных при помощи методов машинного обучения. Использованная база данных была сформирована на основе материалов, находящихся в легальном открытом доступе. Выполнен сравнительный анализ моделей машинного обучения, предназначенных для решения поставленной задачи методами логической регрессии, дерева решений, «случайного леса», а также наивного байесовского классификатора. Сравнивалась результативность перечисленных моделей и двух нейросетевых моделей, включая глубокую нейронную сеть. Посредством всех моделей был выполнен подбор оптимальных параметров путём полного перебора различных комбинаций моделей. Каждый из классификаторов улучшался с применением механизмов стандартизации и снижения размерности методом главных компонент.

Проведенный эксперимент включал создание набора данных, анализ и предварительную обработку данных, обучение и тестирование моделей, их сравнение по различным метрикам качества. По итогам сравнительного анализа моделей наилучший результат продемонстрировала модель нейронной сети, обеспечившая наилучшие показатели по всем метрикам оценки.

Полученные результаты показывают: современные алгоритмы, включая глубокие нейронные сети, эффективно работают в области прогнозирования кредитного риска. Поэтому они могут быть использованы в качестве средств автоматизации при обработке кредитных заявок как больших данных. Перспективой дальнейших исследований является совершенствование итогового результата, включая использование более сложных глубоких нейросетевых моделей, основанных, в частности, на рекуррентных и сверточных нейронных сетях.

<sup>15</sup> Составлено автором.

## Список литературы

1. *Lessmann S., Baesens B., Hsin-Vonn Seow, Thomas Lyn C.* Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research // *European Journal of Operational Research*. – 2015. – Vol. 247, Iss. 1. – P. 124–136. – DOI 10.1016/j.ejor.2015.05.030.
2. *Liu R.* Machine learning approaches to predict default of credit card clients // *Modern Economy*. – 2018. – No. 9. – P. 1828–1838. – DOI 10.4236/me.2018.911115.
3. *Thomas L., Crook J., Edelman D.* *Credit Scoring and Its Applications (Mathematics in Industry)*. SIAM-Society for Industrial & Applied Mathematics. – 2nd Revised edition. – 2017. – 387 p.
4. *Shan Q., Nilsson M.* Credit risk analysis with machine learning techniques in peer-to-peer lending market. – Stockholm University, 2018. – 42 p. – URL: <https://www.diva-portal.org/smash/get/diva2:1375762/FULLTEXT01.pdf> (дата обращения: 28.06.2023). – Текст: электронный.
5. *Курнищикова М.В.* Разработка алгоритма системы оценки кредитного риска банковского сектора в рамках подхода макропруденциального стресс-тестирования // *Digital*. – 2021. – Т. 2, № 2. – URL: <https://cyberleninka.ru/article/n/razrabotka-algoritma-sistemy-otsenki-kreditnogo-riska-bankovskogo-sektora-v-ramkah-podhoda-makroprudentsialnogo-stress> (дата обращения: 04.10.2023). – Текст: электронный.
6. *Ning W.* Application of big data and artificial intelligence technology in computer network // *Electronic technology and software engineering*. – 2019. – No. 08. – P. 12. – URL: [https://www.researchgate.net/publication/347635829\\_Research\\_on\\_the\\_Application\\_of\\_Big\\_Data\\_and\\_Artificial\\_Intelligence\\_Technology\\_in\\_Computer\\_Network](https://www.researchgate.net/publication/347635829_Research_on_the_Application_of_Big_Data_and_Artificial_Intelligence_Technology_in_Computer_Network) (дата обращения: 23.08.2023). – Текст: электронный.
7. *Rabby M. et al.* Stacked LSTM Based Deep Recurrent Neural Network with Kalman Smoothing for Blood Glucose Prediction. – University of Louisiana at Lafayette. – 2021. – URL: <https://arxiv.org/pdf/2101.06850.pdf> (дата обращения: 18.09.2023). – Текст: электронный.
8. *Ting G.* Application of artificial intelligence in computer network technology in the age of big data // *Electronic technology and software engineering*. – 2019. – No. 01. – P. 6. – URL: [https://www.researchgate.net/publication/342204474\\_Application\\_of\\_artificial\\_intelligence\\_in\\_computer\\_network\\_technology\\_under\\_the\\_background\\_of\\_big\\_data\\_era](https://www.researchgate.net/publication/342204474_Application_of_artificial_intelligence_in_computer_network_technology_under_the_background_of_big_data_era) (дата обращения: 08.08.2023). – Текст: электронный.

## References

1. *Lessmann S., Baesens B., Hsin-Vonn Seow, Thomas Lyn C.* Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research // *European Journal of Operational Research*. – 2015. – Vol. 247, Iss. 1. – P. 124–136. – DOI 10.1016/j.ejor.2015.05.030.
2. *Liu R.* Machine learning approaches to predict default of credit card clients // *Modern Economy*. – 2018. – No. 9. – P. 1828–1838. – DOI 10.4236/me.2018.911115.
3. *Thomas L., Crook J., Edelman D.* *Credit Scoring and Its Applications (Mathematics in Industry)*. SIAM-Society for Industrial & Applied Mathematics. – 2nd Revised edition. – 2017. – 387 p.
4. *Shan Q., Nilsson M.* Credit risk analysis with machine learning techniques in peer-to-peer lending market. – Stockholm University, 2018. – 42 p. – URL: <https://www.diva-portal.org/smash/get/diva2:1375762/FULLTEXT01.pdf> (data obrashcheniya: 28.06.2023). – Tekst: elektronnyj.
5. *Kirpishchikova M.V.* Razrabotka algoritma sistemy ocenki kreditnogo riska bankovskogo sektora v ramkah podhoda makroprudencial'nogo stress-testirovaniya // *Digital*. – 2021. – Т. 2, № 2. – URL: <https://cyberleninka.ru/article/n/razrabotka-algoritma-sistemy-otsenki-kreditnogo-riska-bankovskogo-sektora-v-ramkah-podhoda-makroprudentsialnogo-stress> (data obrashcheniya: 04.10.2023). – Tekst: elektronnyj.
6. *Ning W.* Application of big data and artificial intelligence technology in computer network // *Electronic technology and software engineering*. – 2019. – No. 08. – P. 12. – URL: [https://www.researchgate.net/publication/347635829\\_Research\\_on\\_the\\_Application\\_of\\_Big\\_Data\\_and\\_Artificial\\_Intelligence\\_Technology\\_in\\_Computer\\_Network](https://www.researchgate.net/publication/347635829_Research_on_the_Application_of_Big_Data_and_Artificial_Intelligence_Technology_in_Computer_Network) (data obrashcheniya: 23.08.2023). – Tekst: elektronnyj.
7. *Rabby M. et al.* Stacked LSTM Based Deep Recurrent Neural Network with Kalman Smoothing for Blood Glucose Prediction. – University of Louisiana at Lafayette. – 2021. – URL: <https://arxiv.org/pdf/2101.06850.pdf> (data obrashcheniya: 18.09.2023). – Tekst: elektronnyj.

8. *Ting G.* Application of artificial intelligence in computer network technology in the age of big data // Electronic technology and software engineering. – 2019. – No. 01. – P. 6. – URL: [https://www.researchgate.net/publication/342204474\\_Application\\_of\\_artificial\\_intelligence\\_in\\_computer\\_network\\_technology\\_under\\_the\\_background\\_of\\_big\\_data\\_era](https://www.researchgate.net/publication/342204474_Application_of_artificial_intelligence_in_computer_network_technology_under_the_background_of_big_data_era) (data obrashcheniya: 08.08.2023). – Tekst: elektronnyj.