

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего  
образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ  
Н.Г.ЧЕРНЫШЕВСКОГО»**

Кафедра Дифференциальных уравнений и математической экономики

**Анализ кредитного риска заемщиков методами машинного обучения**

**АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

студента 4 курса 441 группы  
направления 09.03.03 Прикладная информатика  
механико-математического факультета  
Гаврилова Максима Михайловича

Научный руководитель

профессор, д.э.н, профессор

\_\_\_\_\_

В.А.Балаш

Заведующий кафедрой

зав. кафедрой, д.ф-м.н, профессор

\_\_\_\_\_

С.И.Дудов

Саратов 2023 г.

## ВВЕДЕНИЕ

С развитием кредитной системы и использованием цифровых технологий для автоматизации процессов кредитования стало возможным собирать и обрабатывать большое количество данных о заемщиках. Однако оценка кредитоспособности всё ещё является трудной задачей, требующей точности и своевременности, чтобы снизить вероятность возникновения кредитных рисков.

Для успешного управления рисками и принятия экономически обоснованных кредитных решений необходимы анализ и оценка кредитного риска.

В последние годы научное сообщество и бизнес-сообщество обратили внимание на возможность применения методов машинного обучения в задаче анализа кредитного риска. Целью исследования является разработка программного обеспечения задачей которого является раскрытие возможностей применения методов машинного обучения для анализа кредитного риска заемщиков.

В работе будет проведен анализ существующих методов анализа кредитного риска, их преимущества и недостатки по сравнению с применением методов машинного обучения.

Основные задачи исследования:

1. Изучить и проанализировать существующие методы оценки кредитного риска заемщиков, их преимущества и недостатки.

2. Проанализировать возможности применения методов машинного обучения для оценки кредитного риска заемщиков по сравнению с традиционными методами.

3. Предложить метод обработки данных для создания модели анализа кредитного риска заемщиков, основанный на методах машинного обучения.

4. Разработать программное обеспечение для применения рассмотренных методов машинного обучения и убедиться в его работоспособности при решении задач анализа кредитного риска.

Основной практический интерес, который может представлять данная работа, заключается в создании модели анализа кредитного риска, основанной на методах машинного обучения. Это может помочь повысить точность расчета кредитного риска и снизить вероятность возникновения неблагоприятных событий.

# **1. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ АНАЛИЗА КРЕДИТНОГО РИСКА**

## **1.1 Анализ предметной области**

Кредитный риск — это вероятность того, что заемщик не в состоянии будет выплатить ссуду в срок. Адекватное определение кредитного риска является важным шагом в процессе оценки кредитоспособности заемщика и анализе рисков, связанных с кредитованием.

Существует множество методов и инструментов для анализа кредитного риска. К некоторым из них относятся:

1. Статистический анализ.
2. Экспертные оценки..
3. Финансовое моделирование.
4. Использование методов машинного обучения и анализа данных.

Методы машинного обучения (МО) играют важную роль в задачах анализа кредитного риска. Они позволяют определить связи между различными факторами, влияющими на вероятность дефолта заемщика, и построить модели для прогнозирования кредитного риска.

1. Классификация
2. Кластеризация
3. Анализ временных рядов
4. Ассоциативные правила
5. Регрессионный анализ

Комбинация различных методов машинного обучения позволяет создавать более точные и надежные модели для анализа кредитного риска. В зависимости от конкретных задач и доступных данных выбираются наиболее подходящие методы.

Среди инструментов, которые применяют для анализа кредитного риска, можно отметить:

1. Кредитные скоринги
2. Кредитные рейтинги

3. Кредитный мониторинг
4. Моделирование вероятности
5. Кредитные комитеты

В целом, все перечисленные методы могут быть эффективными, в зависимости от специфики задачи и требований банка или другой финансовой организации

## **1.2 Обзор программного обеспечения для анализа кредитного риска**

Существует множество программного обеспечения для анализа кредитного риска. Ниже приведены некоторые из наиболее широко используемых:

1. SAS Credit Scoring -.
2. Moody's Analytics RiskCalc
3. Fair Isaac Corporation (FICO)
4. RiskMaster by Algorithmics
5. Credit Risk Navigator by Bloomberg

## **1.3 Сравнительный анализ программного обеспечения для анализа кредитного риска**

На основе проведенного обзора программного обеспечения для анализа кредитных рисков были выделены критерии, которые позволяют провести сравнительный анализ данных приложений. А результаты показаны в таблице.

По строкам таблицы указаны критерии оценки программного обеспечения для анализа кредитных рисков.

Исходя из сравнительного анализа следует, что несмотря на

большие возможности одна копия ПО является весьма дорогостоящей и это говорит о том, что разработка собственного ПО весьма актуально.

## **2. РАЗРАБОТКА ПРОЕКТА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ АНАЛИЗА КРЕДИТНОГО РИСКА**

### **2.1 Анализ целевой аудитории**

Целевой аудиторией данной системы являются финансовые учреждения, такие как банки и кредитные организации, которые выдают кредиты. Также могут быть заинтересованы регуляторные органы, которые контролируют деятельность финансовых учреждений. Пользователи системы могут включать в себя менеджеров по кредитованию, аналитиков, руководителей компании и других специалистов, ответственных за выдачу кредитов и оценку кредитного риска.

### **2.2 Разработка функциональных требований для программного обеспечения для анализа кредитного риска**

Функциональные требования (functional requirements) — описание требуемого поведения системы в определенных условиях.

Функциональные требования определяют, каким должно быть поведение продукта в тех или иных условиях. Они определяют, что разработчики должны создать, чтобы пользователи смогли выполнить свои задачи (пользовательские требования) в рамках бизнес-требований. В результате проведения сравнительного анализа аналогичного программного обеспечения были выделены следующие функциональные требования для разрабатываемого приложения:

Для иллюстрации функциональных требований разработана диаграмма прецедентов (

Для назначения сущностей диаграммы прецедентов построена таблица

## **2.2 Разработка шаблонов интерфейса для программного обеспечения для анализа кредитного риска**

Графический интерфейс разрабатываемого приложения предполагает 4 основных экрана: «Главное Меню», «Анкета» «Результат анализа», «Отправка отчета».

Дизайн графического интерфейса, выполнен в минималистическом стиле и холодных тонах.

## **3. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИНСТРУМЕНТОВ РАЗРАБОТКИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ АНАЛИЗА КРЕДИТНОГО РИСКА**

### **3.1 Обзор и сравнительный анализ языков программирования разработки программного обеспечения для анализа кредитного риска**

Python

C++

Java

R

Сравнивая эти языки программирования, можно отметить, что Python и R обладают более простым синтаксисом и имеют множество библиотек для работы с данными и анализа рисков.

На основе проведенного обзора языков программирования были выделены критерии, которые позволяют провести сравнительный анализ данных приложений. Результаты сравнительного анализа представлены в таблице

### **3.2 Обзор и сравнительный анализ сред программирования**

Среда разработки (IDE) — это интегрированное программное обеспечение, которое упрощает процесс создания, отладки и

тестирования программного кода. Ниже приведен обзор некоторых популярных сред разработки.

1. Visual Studio
2. IntelliJ IDEA
3. Eclipse
4. PyCharm
5. NetBeans

На основе проведенного обзора языков сред разработки были выделены критерии, которые позволяют провести сравнительный анализ данных приложений. Результаты сравнительного анализа представлены в таблице 4.

Независимо от выбора среды разработки, важно выбрать такую, которая подходит для нужных задач и языков программирования, а также обладает мощными инструментами для тестирования, отладки и оптимизации кода.

### **3.3 Обзор и сравнительный анализ библиотек высокоуровневого языка программирования Python для машинного обучения**

Ниже приводится обзор и сравнительный анализ наиболее популярных и мощных из них.

1. Scikit-learn
2. TensorFlow
3. PyTorch
4. CatBoost
5. Chainer

библиотек имеет свои преимущества и недостатки, и в конечном итоге правильный выбор зависит от требований и задач, которые должны быть выполнены. Scikit-learn является наиболее популярной библиотекой.

## **4. РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ АНАЛИЗА КРЕДИТНОГО РИСКА**

### **4.1 Разработка алгоритма для анализа кредитного риска**

Логистическая регрессия — это алгоритм машинного обучения, который используется для классификации данных.

Математическая постановка задачи логистической регрессии была введена

Так же далее был поэтапный разбор функционального кода.

### **4.2 Характеристика набора данных для алгоритма для анализа кредитного риска**

Исходный набор данных содержит 1000 записей с 20 категориальными / символическими атрибутами, подготовленный профессором. Hofmann. В этом наборе данных каждая запись представляет человека, который берет кредит в банке. Каждый человек классифицируется как хороший или плохой кредитный риск в соответствии с набором атрибутов.

### **4.3 Разработка графического интерфейса программного обеспечения для анализа кредитного риска**

В данном разделе происходит Разработка графического интерфейса программного обеспечения для анализа кредитного риска

### **4.4 Разработка модуля отчетов программного обеспечения для анализа кредитного риска**

В данном разделе рассматривается модуль отчета, программного обеспечения и его подробное рассмотрению

## ЗАКЛЮЧЕНИЕ

На основе анализа существующих методов оценки кредитного риска заемщиков, было выявлена необходимость их усовершенствования, так как они не всегда дают точную информацию об оценке риска и могут быть неэффективными в определенных условиях.

Для решения этой проблемы были рассмотрены методы машинного обучения, такие как предиктивное моделирование, классификация и кластеризация данных, нейронные сети и другие. Эти методы показали большой потенциал для оценки кредитного риска заемщиков, как по точности, так и по скорости обработки данных.

Для создания модели анализа кредитного риска заемщиков, основанной на методах машинного обучения, был разработан метод обработки данных. Этот метод был применен для выборки данных, которые были получены из реальной банковской системы, для создания моделей оценки кредитного риска. Результаты апробирования показали, что данная модель даёт более точные результаты, чем традиционные методы анализа кредитного риска.

Для реализации созданной модели было разработано программное обеспечение, которое демонстрировало возможности использования методов машинного обучения в финансовой сфере. Программное обеспечение позволяет быстро обрабатывать большие объемы данных и проводить анализ кредитного риска более точно.

В результате исследования был разработан подход, который можно использовать банками и финансовыми учреждениями для улучшения своих методов анализа кредитного риска. Результаты исследования могут помочь улучшить процессы выдачи кредитов и повысить эффективность банковской сферы в целом.