

Федеральное государственное автономное
образовательное учреждение высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт математики и фундаментальной информатики
Базовая кафедра вычислительных и информационных технологий

УТВЕРЖДАЮ
Заведующий кафедрой
_____ / В.В. Шайдуров

«__» _____ 2022 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

**ИССЛЕДОВАНИЕ МЕТОДОВ РЕШЕНИЯ ЗАДАЧИ
АВТОМАТИЧЕСКОГО КРЕДИТНОГО СКОРИНГА ДЛЯ АНАЛИЗА
ПЛАТЕЖЕСПОСОБНОСТИ КЛИЕНТОВ БАНКА**

Направление 02.04.01 Математика и компьютерные науки

Магистерская программа 02.04.01.02 Вычислительная математика

Руководитель _____ кандидат физико-
математических И.В. Баранова
наук, доцент

Выпускник _____ Т.И. Седельников

Нормоконтролер _____ Т.Н. Шипина

Красноярск 2022

АННОТАЦИЯ

Целью магистерской диссертации является исследование методов оценки кредитного риска и разработка экспертной системы принятия решения о кредитовании.

В основе решения задачи банковского скоринга лежат методы машинного обучения: линейная регрессия, логистическая регрессия, деревья решений, k-средних.

В результате исследования поставлена и решена практическая задача банковского скоринга. Получены условия, при которых достигается максимальная точность предсказания.

Ключевые слова: машинное обучение, банковский скоринг, вероятность, регрессия, метод k-средних.

ANNOTATION

The purpose of the research of the master's thesis is the method of loan risk assessment and the development of an expert system for making a decision on lending.

The solution of the banking scoring problem is based on machine learning methods: linear regression, logistic regression, decision trees, k-means.

As a result of the study, the practical problem of bank scoring was posed and solved. Conditions are obtained under which prediction farsightedness is detected.

Keywords: machine learning, bank scoring, probability, regression, k-means method.

СОДЕРЖАНИЕ

Введение.....	3
1 Основные понятия.....	6
2 Постановка задачи скоринга.....	9
2.1 Типы скоринговых систем.....	9
2.2 Скоринг заявок.....	10
2.3 Характеристики заёмщиков для задачи скоринга.....	12
2.4 Математическая постановка.....	14
3 Методы решения задачи скоринга.....	17
3.1 Точность решения.....	18
3.2 Линейная регрессия.....	20
3.3 Логистическая регрессия.....	21
3.4 Деревья решений.....	23
3.5 Метод k-mean (k-средних).....	26
4 Решение практической задачи.....	30
4.1 Описание статистики.....	31
4.1.1 CFS – отбор признаков на основе корреляции.....	34
5 Описание программного обеспечения.....	35
5.1 Сравнение методов решения задачи.....	40
Заключение.....	44
Список использованных источников.....	45
Приложения.....	47

ВВЕДЕНИЕ

Кредит – это финансовые обязательства двух сторон, одна из которых – банк – предоставляет наличные или другие ресурсы (кредитные карты, ипотеки, рефинансирование и т.д.), а вторая – заемщик – обещает вернуть их согласно принципам срочности, платности и возвратности.

Кредитование, с точки зрения клиента банка (заемщика) – это получение денег в долг на индивидуальных условиях, зафиксированных в договоре. В договоре указывается процентная ставка, срок на который предоставляются деньги, дата и размер ежемесячного платежа.

На фоне постоянно растущей конкуренции в банковской отрасли, в связи с постоянным появлением новых банковских продуктов или открытием новых банков, основополагающим критерием для развития и конкурентоспособности банков становятся такие факторы, как уменьшение времени на принятие решения о предоставлении какого-либо банковского продукта, снижение требований к его обеспечению и уменьшение количества предоставляемой информации заемщиками при подаче заявки на кредит.

Для улучшения и упрощения процесса проверки и верификации заемщиков необходима эффективная автоматизация бизнес-процессов, которая в долгосрочной перспективе позволит уменьшить долю рискованных кредитных активов банка. Это возможно благодаря наличию в банковских структурах специализированных скоринговых систем, основанных на вычислениях и предсказаниях результата с помощью методов машинного обучения.

Кредитный скоринг – используемая банками структура оценки кредитоспособности (кредитных рисков) физических лиц, в основе которой лежат численные статистические методы. Скоринг является упрощённой системой анализа характеристик заёмщика, что позволяет снизить требования к квалификации кредитного менеджера, который занимается обработкой заявок на кредитный продукт, и увеличить скорость их рассмотрения. Развитие

компьютерных систем и технологий привело к идее автоматизации скоринга, т.е. созданию экспертной системы, способной анализировать кредитоспособность заёмщика на основании данных о нём и выдавать рекомендацию о выдаче кредита или отказе в нём.

В магистерской диссертации объектом исследования является задача кредитного скоринга, то есть система оценки кредитоспособности.

Предметом исследования являются алгоритмы оценки кредитоспособности заёмщика и методы точности предсказаний, обеспечивающие поддержку принятия решения в процессе кредитования.

Целью магистерской диссертации является исследование методов оценки кредитного риска и разработка экспертной системы принятия решения о кредитовании.

Чтобы достичь поставленной цели были поставлены и решены следующие задачи:

- Проанализировать процесс кредитования клиентов;
- Изучить постановку задачи кредитного скоринга и основные алгоритмы решения (то есть алгоритмы прогнозирования платежеспособности клиента);
- Проанализировать представленный набор данных (датасет);
- Провести отбор наиболее важных показателей, описывающих кредитоспособность и благонадежность клиента;
- Разработать программное обеспечение, реализующее работу основных методов решения задачи скоринга;
- Выполнить сравнение рассмотренных алгоритмов и исследовать эффективность прогнозирования платежеспособности;
- Решить практическую задачу банковского скоринга;
- Провести анализ полученных результатов решения практической задачи.

Теоретической значимостью исследования является разработка комплексного подхода к построению и обработке «кредитных» данных, а также применению метода k-средних, ранее не использовавшегося в решениях задачи.

Практическая значимость исследования заключается в реализации теоретических предположений в виде программного обеспечения, которое позволяет снизить расходы и увеличить прибыль коммерческих организаций, занимающихся кредитованием.

В первой главе диссертации приводятся основные понятия из теории статистики, обработки данных, кластерного анализа и машинного обучения.

Во второй главе дается постановка задачи скоринга. Также в этой главе дается информация о существующих типах (системах) скоринга, и указываются особенности задачи.

В третьей главе рассматриваются используемые в работе подходы к решению задачи (линейная регрессия, логистическая регрессия, деревья решений, метод k-средних) и методы определения точности решений (предсказаний).

В четвертой главе описана и решена практическая задача, для которой были произведены вычислительные эксперименты, и создан программный модуль, реализующий работу всех необходимых методов.

В пятом разделе описывается архитектура разработанного программного обеспечения, демонстрируется графический интерфейс и особенности его работы. Также выполняется сравнительный анализ полученных результатов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Чубукова, И. А. Data mining: научное пособие / И. А. Чубукова. – Киев: КНЭУ, 2008. – 326 с.
2. Барсегян, А. А. Технология анализа данных. Data Mining, Visual Mining, Text mining, OLAP / А. А. Барсегян. – Спб.: БХВ, 2004. – 250 с.
3. Воронцов, К. В. Лекции по алгоритмам кластеризации и многомерного шкалирования / К. В. Воронцов. – Москва: МГУ, 2007. – 18 с.
4. Будина, Е. С. Применение системы кредитного скоринга как способ повышения конкурентоспособности кредитной организации / Е. С. Будина. – Москва, Московский государственный университет экономики, статистики и информатики, 2009.
5. Грюнштейн, Джон М. Л. Оптимальное использование статистических методов при построении моделей / Джон М. Л. Грюнштейн; под ред. Элизабет Мейз; пер. с англ. И. М. Тикота; науч. ред. Д. И. Вороненко. – Москва: «Гревцов Паблишер», 2008.
6. Мельникова, А. В. Оптимизация процесса предкредитной обработки. Эффективное принятие решений / А. В. Мельникова. – Москва: «Банковское кредитование», 2007.
7. Малюгин, В.И. Об эффективности статистических алгоритмов кредитного скоринга. / В.И. Малюгин, Н.В. Гринь. – Москва: Банковский вестник – No 31, 2010, 39–46 с.
8. Макконелл, Дж. Основы современных алгоритмов / Дж. Макконелл, пер. с англ. под ред. С. К. Ландо. – Изд. 2-е, перераб. и доп. – М.: Техносфера, 2004. – 366 с.
9. Левитин, А. В. Алгоритмы. Введение в разработку и анализ / А. В. Левитин. – Москва: Вильямс, 2006. – 35–36 с.
10. Захарова, Е.М. Обзор методов многомерной оптимизации / Е. М. Захарова; И. К. Минашина. – Москва: Информационные процессы, 2014. –

Том 14. – № 3. 265– 266 с.

11. Быкова, В. В. Дискретная математика с использованием ЭВМ: учебное пособие / В. В. Быкова. – Красноярск: РИО КрасГУ, 2006, 200 с.
12. Местецкий, Л. М. Математические методы распознавания образов: курс лекций / Л. М. Местецкий. – Москва: МГУ, 2004, 85 с.
13. Соколов, Г. А. Введение в регрессионный анализ и планирование регрессионных экспериментов в экономике / Г. А. Соколов. – Москва: ИНФРА, 2017. – 109 с.
14. Смирнов, В. Д. Методы корреляционно-регрессионного анализа в эконометрических исследованиях: учебное пособие / В. Д. Смирнов, А. В. Белокопытов. – Смоленск: СГСА, 2004. – 104 с.
15. Зализняк, В. Е. Введение в математическое моделирование: учебное пособие / В. Е. Зализняк, О. А. Золотов. – Москва: Юрайт, 2020. – 133.

Министерство науки и высшего образования РФ
Федеральное государственное автономное
образовательное учреждение высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт математики и фундаментальной информатики
Базовая кафедра вычислительных и информационных технологий

УТВЕРЖДАЮ

Заведующий кафедрой
Шайду / В.В. Шайдуров

«16» 06 2022 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

**ИССЛЕДОВАНИЕ МЕТОДОВ РЕШЕНИЯ ЗАДАЧИ
АВТОМАТИЧЕСКОГО КРЕДИТНОГО СКОРИНГА ДЛЯ АНАЛИЗА
ПЛАТЕЖЕСПОСОБНОСТИ КЛИЕНТОВ БАНКА**

Направление 02.04.01 Математика и компьютерные науки

Магистерская программа 02.04.01.02 Вычислительная математика

Руководитель

И.В. Баранова
16.06.2022

кандидат физико-
математических
наук, доцент

И.В. Баранова

Выпускник

Т.И. Седельников
16.06.2022

Т.И. Седельников

Нормоконтролер

Т.Н. Шипина
23.06.22

Т.Н. Шипина

Красноярск 2022