

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

ФАКУЛЬТЕТ КОМПЬЮТЕРНЫХ НАУК
ОСНОВНАЯ ОБРАЗОВАТЕЛЬНАЯ ПРОГРАММА
«ПРИКЛАДНАЯ МАТЕМАТИКА И ИНФОРМАТИКА»

КУРСОВАЯ РАБОТА

Программный проект на тему:
«Нейронные сети на табличных данных на примере
задачи кредитного скоринга»

Выполнила студентка группы БПМИ194 3 курса,
Смирнова Анна Романовна

Руководитель КР:

Преподаватель Воробьева Мария Сергеевна

Оглавление

Аннотация.....	3
Abstract.....	4
Введение	5
Обработка данных	7
Список литературы.....	9

Аннотация

Программный проект на тему: «Нейронные сети на табличных данных на примере задачи кредитного скоринга»

Выполнила: Смирнова А. Р.

Руководитель: Воробьева М. С.

В ходе данной курсовой работы был выполнен анализ качества существующих методов решения задачи кредитного скоринга. Главной целью являлось проанализировать и протестировать существующие методы решения данной задачи, после чего составить подробное описание каждого из методов и провести сравнительный анализ качества. Методы, которые были изучены: Ridge-регрессия, логистическая регрессия, метод опорных векторов, дерево решений, случайный лес.

Ключевые слова: кредитный скоринг, модель, метрика качества, модели машинного обучения, банковская сфера, кредит.

Abstract

Program course work: «Credit scoring. Linear models vs modern data science models»

Student: Smirnova A. R.

Teacher: Vorobeva M. S.

In the course work, the analysis of the quality of existing methods of solving the problem of credit scoring was carried out. The main objective was to analyze and test existing methods of solving this problem. The conclusion included a detailed description of each method and a comparative analysis of quality. Methods that have been studied: Ridge-regression, logistic regression, support vector machine, decision tree, random forest.

Keywords: credit scoring, neural network, machine learning models, quality metric, banking industry, loan.

Введение

Задача кредитного скоринга возникает в ситуации, когда банку необходимо принять решение о выдаче или отказе по кредиту, при этом принимая во внимание множество несвязанных факторов. Одним из решений данной ситуации является субъективное заключение кредитного эксперта, однако человеческий фактор не всегда позволяет учесть все входные данные, потому что с течением времени количества факторов, влияющих на принятие решение о выдаче кредита, растет.

Банковская сфера активно расширяется, и каждой компании необходимо внедрять качественные автоматизированные системы, которые помогут эффективно обеспечивать контроль работы бизнес-процессов. Решить кредитный вопрос активно помогают системы кредитного скоринга, такие как, например, нейронные сети на табличных данных, или другие модели классификации, с помощью которых можно оценить большой набор признаков и учесть предыдущий опыт работы с кредиторами.

Главной **целью** этой работы является самостоятельно применить на табличных данных существующие методы решения задачи кредитного скоринга, после чего составить подробное описание каждого из методов и провести сравнительный анализ качества. **Актуальность** работы подтверждается постоянным совершенствованием банковской системы, и растущей необходимостью работать с большими объемами данных. Численность населения непременно растет, в связи с чем увеличивается и количество предлагаемых банками услуг, в том числе различного рода кредитов, отличных от привычного понимания. Например, появлением лизинга¹ или популярности микрозаймов², которые так же нуждаются в

¹ Лизинг — это долгосрочная аренда определенных объектов собственности (оборудование, машины, сооружения) с погашением задолженности в течение нескольких лет.

² Микрозаймы — это кредиты на небольшие денежные суммы, которые обычно даются на короткий период.

анализе платежеспособности своих клиентов.

Основные задачи, которые предстоит выполнить в ходе работы:

1. Предварительная обработка имеющихся данных.
2. Реализация различных методов машинного обучения (**тут допишу по мере реализации**): линейная регрессия, Lasso регрессия, Ridge-регрессия, Логистическая регрессия с WOE переменными
3. Сравнительный анализ полученных результатов через метрики качества: MSE, RMSE и другие.
4. Формулирование предложений о том, как можно внедрить данные системы для эффективной работы в банковской сфере.

Обработка данных

Линейная регрессия

Одним из самых широко известных и простых алгоритмов машинного обучения является линейная регрессия.

Список литературы

1. Ahamed, S. (б.д.). *Machine learning algorithms can help us to estimate the risk of a financial decision*. Получено из Towards Data Science: <https://towardsdatascience.com/financial-data-analysis-80ba39149126>
2. *Complex System Modelling in Engineering Under Industry*. (22 10 2021 г.). Получено из Hindawi: <https://www.hindawi.com/journals/complexity/2021/9222617/>
3. Essche, P. O. (2018). *Solvay Brussels School*. Получено из Big Data for Credit Scoring: towards the End of Discrimination on the Credit Market? Evidence from Lending Club.
4. *Python_Credit Scoring (ML)_Elastic net regression*. (б.д.). Получено из kaggle: <https://www.kaggle.com/code/aashofteh/python-credit-scoring-ml-elastic-net-regression/data>
5. *Отбор признаков в задачах машинного обучения*. (б.д.). Получено из Хабр: <https://habr.com/ru/post/550978/>

