

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра математического моделирования и анализа данных

Аннотация к магистерской диссертации

**«Разработка моделей кредитного скоринга заёмщиков коммерческих
банков с использованием методов машинного обучения »**

ШЭНЬ Сяююй

Научный руководитель – кандидат доктор экономических наук,
кандидат физ.-мат. наук, профессор
Малюгин В.И.

Минск, 2025

Реферат

Магистерская диссертация, 47 страниц, 10 таблиц, 15 рисунков, 36 источников, 41 формула.

Ключевые слова: МАШИННОЕ ОБУЧЕНИЕ, КРЕДИТНЫЙ СКОРИНГ, АНСАМБЛЕВОЕ ОБУЧЕНИЕ, МОДЕЛЬ STACKING, ДИСБАЛАНС КЛАССОВ, МЕТОД BORDLINE-SMOTE, МЕТОД LIME.

Объектом исследования являются надёжность заёмщиков коммерческих банков на основе данных о их финансовом поведении.

Предметом исследования является методы машинного обучения для построения моделей кредитного скоринга, включая одиночные и ансамблевые алгоритмы.

Целью работы является построение модели кредитного скоринга для оценки надежности заемщиков с использованием наиболее часто применяемых методов машинного обучения, в том числе ансамблевого метода Stacking и улучшение интерпретации сложных моделей машинного обучения с использованием метода LIME.

В данной работе разработаны модели кредитного скоринга с использованием 9 алгоритмов машинного обучения, включая логистическая регрессия, SVM, XGBoost и др. После сравнения их метрик эффективности, предлагается модель кредитного скоринга на основе алгоритма Stacking, который использует интеграцию восьми алгоритмов, наиболее часто применяемых в задачах кредитного скоринга. Как известно, Stacking целесообразно применять в условиях данных сложной структуры. В работе используемый алгоритм Stacking показал достаточно высокую точность и стабильность в прогнозировании кредитного риска. Экспериментальные результаты на реальных данных показали, что модель кредитного скоринга, использующая ансамблевый алгоритм Stacking превосходит все входящие в него одиночные классификационные алгоритмы по ряду основных оценочных метрик, включая AUC и F1-Score. Это позволяет рассматривать разработанную модель кредитного скоринга как эффективный инструмент для оценки кредитных рисков в коммерческих банках.

Abstract

Master's thesis, 47 pages, 10 tables, 15 figures, 36 sources, 41 formulas.

Keywords: MACHINE LEARNING, CREDIT SCORING, ENSEMBLE LEARNING, STACKING MODEL, CLASS IMBALANCE, BORDERLINE-SMOTE METHOD, LIME METHOD.

The object of the research is the creditworthiness of commercial bank borrowers based on data about their financial behavior.

The subject of the research is machine learning methods for building credit scoring models, including single and ensemble algorithms.

The aim of the work is to develop a credit scoring model for assessing borrower reliability using the most commonly applied machine learning methods, including the ensemble Stacking method, and to improve the interpretation of complex machine learning models using the LIME method.

In this work, credit scoring models were developed using 9 machine learning algorithms, including Logistic Regression, SVM, XGBoost, etc. After comparing their performance metrics, a credit scoring model based on the Stacking algorithm is proposed. This Stacking model integrates eight algorithms most frequently used in credit scoring tasks. As is known, Stacking is advisable to use with data of complex structure. In this work, the employed Stacking algorithm demonstrated sufficiently high accuracy and stability in predicting credit risk. Experimental results on real data showed that the credit scoring model using the Stacking ensemble algorithm outperforms all its constituent single classification algorithms across several key evaluation metrics, including AUC and F1-Score. This allows the developed credit scoring model to be considered an effective tool for assessing credit risks in commercial banks.