

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра дифференциальных уравнений и системного анализа

Аннотация к дипломной работе

Математические методы анализа данных в банкинге

Гоман Екатерина Юрьевна

Научный руководитель:
доктор физ.-мат. наук,
профессор В. И. Громак

В дипломной работе 41 страница, 14 иллюстраций, 1 таблица, 10 источников, 3 приложения.

РИСК-АНАЛИТИКА, КРЕДИТНЫЙ СКОРИНГ, МОДЕЛИ PROBABILITY OF DEFAULT, КРИВЫЕ RATE, ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ, БУСТИНГ

Объектом исследования дипломной работы являются методы решения задач кредитного скоринга.

Целью дипломной работы является разработка и анализ моделей вероятности дефолта (Probability of Default) и построение кривых rate для прогнозирования кредитного риска.

Для достижения поставленной цели были использованы методы логистической регрессии и бустинга, а также подходы к линеаризации признаков и анализу их информативности. В работе применялись статистические методы, машинное обучение и инструменты визуализации данных.

В дипломной работе получены следующие результаты:

1. Разработан функционал для линеаризации признаков. Линеаризовано 5 признаков на двух сегментах, что прибавило 4% к ROC AUC модели.
2. Составлен конвейер отбора признаков. Модель была упрощена в 4 раза без потери ROC AUC.
3. Реализованы функции для построения кривых rate, выявлены скрытые закономерности в данных.

Дипломная работа является завершённой, поставленные задачи решены в полной мере, присутствует возможность дальнейшего развития исследований.

Дипломная работа выполнена автором самостоятельно.

The thesis consists of 41 pages, 14 illustrations, 1 table, 10 references, and 3 appendices.

RISK ANALYTICS, CREDIT SCORING, PROBABILITY OF DEFAULT MODELS, RATE CURVES, LOGISTIC REGRESSION, BOOSTING

The object of this thesis is methods for solving credit scoring tasks.

The aim of the thesis is to develop and analyze probability of default (PD) models and construct rate curves for credit risk prediction.

To achieve this goal, methods of logistic regression and boosting were used, as well as approaches to feature linearization and informativeness analysis. The study applied statistical methods, machine learning, and data visualization tools.

The following results were obtained in the thesis:

1. Functionality for feature linearization was developed. Five features were linearized in two segments, which increased the model's ROC AUC by 4%.
2. A feature selection pipeline was designed. The model was simplified fourfold without loss of ROC AUC.
3. Functions for constructing rate curves were implemented, and hidden patterns in the data were revealed.

The thesis project is complete, all tasks have been successfully done, there is a possibility for further research and development.

The thesis project was done solely by the author.