

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ**  
**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**  
**ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ**  
**Кафедра математического моделирования и анализа данных**

Аннотация к дипломной работе  
**«Разработка модели кредитного скоринга с использованием  
алгоритмов машинного обучения»**

Пицко Дарья Анатольевна

Научный руководитель – кандидат физ.-мат. наук, доцент В.И.Малюгин

2017

## АННОТАЦИЯ

*Дипломная работа, 57 с., 35 рис., 53 табл., 19 источника.*

МАШИННОЕ ОБУЧЕНИЕ, ЗАДАЧА КЛАССИФИКАЦИИ, АЛГОРИТМ ДЕРЕВА РЕШЕНИЙ C4.5, АЛГОРИТМ ДЕРЕВА РЕШЕНИЙ CART, ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ, АЛГОРИТМ КЛАССИФИКАЦИИ СЛУЧАЙНЫЙ ЛЕС, КРЕДИТНЫЙ СКОРИНГ, АНАЛИЗ КАЧЕСТВА МОДЕЛИ КРЕДИТНОГО СКОРИНГА.

*Объектом исследования* является задача кредитного скоринга, проблемы классификации заёмщиков и предсказания соответствующих классов с помощью методов машинного обучения на основе различных моделей.

*Цель работы* – разработка и исследование моделей и алгоритмов для классификации заёмщиков банка и предсказания классов заёмщиков на анкетных данных и данных из социальной сети.

За время работы были решены следующие задачи:

- дан общий анализ проблемы разработки моделей кредитного скоринга на основе методов машинного обучения;
- для задачи кредитного скоринга, рассматриваемой как задача статистической классификации, описаны используемые в работе алгоритмы машинного обучения логистической регрессии, C4.5, CART, Random Forest;
- построены модели кредитного скоринга с помощью алгоритмов C4.5, CART, Random Forest на языке программирования Python для двух наборов данных: данные, которые собираются банком с помощью анкеты и уже имеющейся кредитной истории клиента и данные, извлекаемые из социальных сетей;
- в процессе построения моделей кредитного скоринга изучены возможности пакетов языка программирования Python по подготовке данных numpy, pandas, matplotlib а также пакета sklearn по построению предикативных моделей для задачи классификации;
- проведен сравнительный анализ моделей кредитного скоринга для различных алгоритмов классификации.

Полученные результаты могут использоваться в сфере кредитования для совершенствования моделей кредитного скоринга путем добавления признаков, полученных из социальных сетей к основной анкете.

## ABSTRACT

*Diploma thesis, 57 p., 35 pic., 53 tables, 19 sources.*

MACHINE LEARNING, CLASSIFICATION, DECISION TREE C4.5, DECISION TREE CART, LOGISTIC REGRESSION ALGORITHM, RANDOM FOREST ALGORITHM, CREDIT SCORING, CREDIT SCORING MODEL'S QUALITY.

*Object of research* – credit scoring models, problems of the classification of borrowers and predictions of borrowers' classes using machine learning methods based on different models.

*Purpose of research* – development and research of the models and algorithms for classification of borrowers and predictions of borrowers' classes based on data from bank and data from social networks.

During the work, the following tasks were solved:

- general analysis of the problem of developing models of credit scoring based on machine learning methods is given;
- for the problem of credit scoring, considered as a problem of statistical classification, the algorithms of machine learning for logistic regression, C4.5, CART, Random Forest used in the work are described;
- models of credit scoring using algorithms C4.5, CART, Random Forest in the programming language Python for two sets of data: data that the bank collects using the questionnaire and the client's existing credit history and data extracted from social networks were built;
- in the process of building credit scoring models, the capabilities of the Python programming language were studied packages for preparing data numpy, pandas, matplotlib and also the sklearn package for constructing predicative models for the classification problem;
- comparative analysis of credit scoring models for different classification algorithms was carried out.
- The results obtained can be used in the sphere of credit for improving models of credit scoring by adding signs obtained from social networks to the main questionnaire.

The obtained results can be used in the sphere of credit for improving models of credit scoring by adding features from social networks to the main questionnaire.