

Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»

Факультет інформаційних технологій
(факультет)

Кафедра системного аналізу і управління
(повна назва)

ПОЯСНЮВАЛЬНА ЗАПИСКА
Кваліфікаційної роботи ступеню магістра
(назва освітньо-кваліфікаційного рівня)

студентки М'ясоїд Тетяни Олексіївни
академічної групи 124М-19-1
спеціалізації _____
за освітньо-професійною програмою 124 Системний аналіз
на тему: «Використання методів інтелектуального аналізу даних для підвищення
точності класифікації споживачів кредитів»

Керівники	Прізвище, ініціали	Оцінка за шкалою		Підпис
		рейтинговою	інституційною	
кваліфікаційної роботи	<i>Д.м.-ф.н., проф. О. П. Купенко</i>			
розділів:	2			
<i>Інформаційно- аналітичний</i>	<i>Д.м.-ф.н., проф. О. П. Купенко</i>			
<i>Спеціальний</i>	<i>Д.м.-ф.н., проф. О. П. Купенко</i>			
Рецензент	<i>к.т.н., доц. Гуліна І.Г.</i>			
Нормоконтролер	<i>к.т.н., доц. Малієнко А.В.</i>			

Дніпро
2020

РЕФЕРАТ

Пояснювальна записка: 97 с., 40 рис., 1 табл., 3 додатки, 39 джерел.

Скорингові системи – популярний в банківській сфері інструмент оцінки клієнтів, в основі якого закладені статистичні методи. Такі системи дозволяють знизити витрати і мінімізувати операційний ризик за рахунок автоматизації прийняття рішення про надання кредиту.

Об’єкт дослідження: анкетні дані споживачів кредитів для задачі кредитного скорингу.

Предмет дослідження: алгоритм машинного навчання XGBoost, заснований на дереві пошуку рішень, який використовує фреймворк градієнтного бустинга.

Мета дослідження: дослідження якості класифікації позичальників в задачі кредитного скорингу за допомогою моделі алгоритму машинного навчання XGBoost.

Методи дослідження та апаратура: методи інтелектуального аналізу та класифікації даних, мова програмування Python.

Економічний ефект від реалізації результатів дослідження очікується позитивним за рахунок використання суб’єктами господарювання алгоритму препроцесінгу даних та побудови релевантної моделі алгоритму машинного навчання XGBoost для оптимізації їх діяльності.

В інформаційно-аналітичному розділі наведено визначення кредитного скорингу, описано основні алгоритми для вирішення задачі кредитного скорингу, проведено аналіз об’єкту дослідження, розглянуто послідовність препроцесінгу даних та методів вибору оптимального набору інформативних ознак. Описано можливості мови програмування Python для інтелектуального аналізу даних.

У спеціальному розділі виконано аналіз та підготовку вихідних даних, створено модель алгоритму XGBoost та відібрано найбільш інформативні ознаки, проведено аналіз параметрів моделі та відібрано ті, що максимально покращують результати прогнозу; представлено результати розв’язку задачі кредитного скорингу.

Практична цінність роботи полягає у розробці детального алгоритму препроцесінгу даних, побудові релевантної моделі алгоритму машинного навчання XGBoost та дослідженні її на адекватність для ефективного прогнозу кредитоспроможності позичальників з метою зниження ризиків для банківської установи при видачі кредиту.

Ключові слова: КРЕДИТНИЙ СКОРИНГ, КРЕДИТОСПРОМОЖНІСТЬ, ДЕРЕВО КЛАСИФІКАЦІЇ, ПРЕПРОЦЕСІНГ, КАТЕГОРІАЛЬНІ ДАНІ, МАШИННЕ НАВЧАННЯ, PYTHON, XGBOOST.

THE ABSTRACT

Explanatory note: 97 pages, 40 pic., 1 table, 3 applications, 39 sources.

Scoring systems are a popular customer assessment tool in the banking sector, based on statistical methods. Such systems can reduce costs and minimize operational risk by automating the decision to grant a loan.

Object of research: personal data of credit consumers for the task of credit scoring.

Subject of research: XGBoost machine learning algorithm, based on a solution search tree, that uses a gradient boosting framework.

Objective: researching of the quality of classification of borrowers in the problem credit scoring models using machine learning algorithm XGBoost.

Research methods: methods of data mining and classification, programming language – Python.

The *economic effect* of the implementation of the results of the study is expected to be positive due to the fact that businesses use the created algorithm for data preprocessing and build a relevant model of the machine learning algorithm XGBoost to optimize their activities. *The practical significance of the results:* is that they can be successfully used in the real management of complex modern systems.

The *information and analytical section* defines credit scoring, describes the main algorithms for solving the problem of credit scoring, analyzes the object of study, considers the sequence of data preprocessing and methods of selecting the optimal set of informative features. The capabilities of the Python programming language for data mining are described.

In a *special section* the analysis and preparation of initial data is performed, the model of the XGBoost algorithm is created and the most informative features are selected, the analysis of model parameters is carried out and the ones that improve the forecast results are selected; the results of solving the problem of credit scoring are presented.

The *practical value* of the work is to develop a detailed algorithm for data preprocessing, build a relevant model of machine learning algorithm XGBoost and study

its adequacy for effective forecasting of solvency of borrowers to reduce risks for the bank when issuing a loan.

Keywords: CREDIT SCORING, SOLVENCY, CLASSIFICATION TREE, PREPROCESSING, CATEGORY DATA, MACHINE LEARNING, PYTHON, XGBOOST.