

Торайғыров университетінің  
ҒЫЛЫМИ ЖУРНАЛ

НАУЧНЫЙ ЖУРНАЛ  
Торайғыров университета

---

# ТОРАЙҒЫРОВ УНИВЕРСИТЕТІНІҢ ХАБАРШЫСЫ

Физика, математика және компьютерлік  
ғылымдар сериясы  
1997 жылдан бастап шығады



## ВЕСТНИК ТОРАЙҒЫРОВ УНИВЕРСИТЕТА

Серия: Физика, математика  
и компьютерные науки  
Издается с 1997 года

---

ISSN 2959-068X

№ 1 (2023)  
Павлодар

**НАУЧНЫЙ ЖУРНАЛ  
ТОРАЙГЫРОВ УНИВЕРСИТЕТА**

**Серия: Физика, математика и компьютерные науки**  
выходит 4 раза в год

---

**СВИДЕТЕЛЬСТВО**

о постановке на переучет периодического печатного издания,  
информационного агентства и сетевого издания  
№ KZ91VPY00046988

выдано

Министерством информации и общественного развития  
Республики Казахстан

**Тематическая направленность**

публикация материалов в области физики, математики,  
механики и информатики

**Подписной индекс – 76208**

<https://doi.org/10.48081/SGQS7560>

---

**Бас редакторы – главный редактор**

Тлеуменов С. К., *ф.-м.г.д., профессор*

Заместитель главного редактора    Испулов Н. А., *ф.-м.г.к., профессор*

Ответственный секретарь            Жумабеков А. Ж., *PhD*

**Редакция алкасы – Редакционная коллегия**

|                    |                                                     |
|--------------------|-----------------------------------------------------|
| Esref Adali,       | <i>PhD докторы, профессор (Турция);</i>             |
| Qadir Abdul,       | <i>PhD докторы, профессор (Пакистан);</i>           |
| Донбаев К. М.,     | <i>ф.-м.г.д., профессор;</i>                        |
| Демкин В. П.,      | <i>ф.-м.г.д., профессор (Российская Федерация);</i> |
| Жумадилаева А. К., | <i>т.г.к., кауымд. профессор;</i>                   |
| Ибраев Н. Х.,      | <i>ф.-м.г.д., профессор;</i>                        |
| Косов В. Н.,       | <i>ф.-м.г.д., профессор;</i>                        |
| Сентова С. М.,     | <i>пед.г.д., профессор;</i>                         |
| Шоканов А. К.,     | <i>ф.-м.г.к., профессор</i>                         |
| Омарова А. Р.,     | <i>технический редактор</i>                         |

---

За достоверность материалов и рекламы ответственность несут авторы и рекламодатели

Редакция оставляет за собой право на отклонение материалов

При использовании материалов журнала ссылка на «Вестник Торайгыров  
университета» обязательна

© Торайгыров университет

**\*D. D. Yessenaman<sup>1</sup>, L. M. Alimzhanova<sup>2</sup>, A. K. Sarbasova<sup>3</sup>**

<sup>1,2,3</sup>Al-Farabi Kazakh National University, Republic of Kazakhstan, Almaty

\*e-mail: [yessenamandiana@gmail.com](mailto:yessenamandiana@gmail.com)

## **MODELING SOFTWARE FOR AN EFFECTIVE BANKING SCORING SYSTEM**

*The article is devoted to the scoring system of lending in second-tier banks. Namely, a decision-making system in which the integration of two scoring models (banking and insurance) leads to an assessment of the reliability of the borrower and the potential risks of the bank, as well as the use of social networks to determine the riskiness of the client. Bank and insurance scoring is compared. The article includes methodologies, processes and data that are used in the development of credit scoring. The advantages of creditworthiness assessment methodologies are described without using the data of insurance companies. Analytical methods are used, receiving data from social networks.*

*Statistical theory is applied to real situations of credit scoring. The article is useful for working people working in the financial sector. These include credit officers, risk managers and analysts, IT-specialists.*

*IT infrastructure, operating procedures, business processes are being studied. An overview of IT tools for the implementation of the scoring system product is given, which increases the efficiency of the scoring system.*

*Keywords: system, scoring, model, risk, income, borrower, lending.*

### **Introduction**

All over the world, automated scoring systems are rightfully considered one of the main methods for assessing the risk of lending. Scoring is a statistical and mathematical tool based on which a retrospective analysis of the credit activity of bank customers takes place, which makes it possible to determine the probability of reliability (or vice versa, unreliability) of a potential borrower [1].

Retail lending in the last 3–4 years has been playing the role of the flagship of the banking sector. Banks active in retail receive hundreds of thousands of applications per month, but due to limited human resources, they cannot analyze each one separately. Therefore, the process is automated and all information

received from the applicant is analyzed using scoring models [2]. To improve the quality of their scoring, STBs use a huge amount of information that they can obtain legally. In addition to databases, which have already become classic sources of information, banks are increasingly turning to alternative criteria for assessing potential borrowers [3].

Scoring technologies are used in various fields. For the first time they were used in the banking sector to assess the solvency of the borrower, later scoring began to be used in other areas of activity. The name scoring comes from the English word «score» (account). Scoring is a system for assessing clients based on statistical methods. This is a computer program where the data of a potential client is entered. As a result, which provides an answer whether to issue a loan or not [4]. There are 4 types of scoring:

Application scoring (translated from English – «scoring of applications, appeals») – an assessment of the borrower's solvency when issuing a loan. This is the most common type of scoring. In this type of scoring, the personal data of the borrower is collected, processed by the program and gives the result: refusal or approval of the loan [5].

Collection scoring – type of scoring at the stage of work with bad loans. The priority actions of bank employees are determined for the return of «outstanding» loans, from warning to transferring the case to a collection agency. It is believed that in the process of such processing, usually 40 % of customers forget to repay the loan.

Behavioral scoring, «scoring of behavior» – assessment of the most likely actions of the borrower. The system predicts changes in the borrower's solvency and adjusts the credit limits set for him [6].

Fraud scoring – statistical assessment of the probability of fraudulent actions of the borrower. With the help of behavioral-scoring, the bank evaluates the behavior of the client over a long period of time.

### **Materials and methods**

The method of applying the scoring system of insurance companies in STBs leads to an assessment of the borrower's reliability and the bank's potential risks. The main factors for risk modeling are:

- the number of overdue payments,
- the frequency of violations,
- mortgage, car loans.

With the help of innovative technologies, a general idea of the driver on the road is compiled using telematic devices. The operation process of the telematics device is shown in Figure 2.



Figure 1 – The operation process of the telematics device

The telematics device is installed in the vehicle's OBD socket. The device records driving performance using an accelerometer and an odometer. The indicators are transmitted for processing via satellite via a cellular communication channel. The data is processed through a scoring model and transferred to the Insurer. Based on the results of scoring, the client is assigned a certain number of points.

The equipment monitors and transmits to the operator an array of information: average speed, tendency to sudden braking and dangerous maneuvers, number of trips, road incidents. Based on the results, the motorist is assigned a score that will be used in lending to the borrower at the bank. In this way, the bank evaluates borrowers as «accident-free» drivers. After making sure that the driver saves his car and prefers not to violate traffic rules, the decision-making system assigns a score, which leads to an increase in the possibility of lending. Using insurance scoring systems, you can determine the level of riskiness of the client. If the client is less risky, then this means the person is more guaranteed to repay the loan [7].

Social networks contain a large amount of information, which makes it possible to obtain data on the behavior of customers, to determine whether a person is risky or not. To conduct scoring based on the client's social networks, tools will be used to verify the borrower and his profile on the social network, parsing social network profiles, obtaining data on user activity through the API, as well as the data storage itself [8].

The key tool for extracting data is the API. It includes a package of classes, procedures, functions, structures and constants for use in external software products offered by the library, service. To access the API, the user of the social network must give permission to collect their personal data. Figure 2 shows the architecture of the solution.

The data center architecture is designed to solve the following tasks:

- Downloading all social network profiles;
- Adding newly created profiles on a regular basis;
- Downloading detailed data (subscriptions, communities, friends, posts) of social networks with a difference of 3 years;

Regular downloading of detailed data of social networks (subscriptions, communities, friends, new wall posts since last upload).



condition, to bankruptcy, non-repayment of debts to the bank. Scoring determines who will receive a loan, the features of the loan issued, and whether the client is a fraudster. Thus, scoring systems allow you to reduce the level of risks without loss of profitability by determining how problematic the bank's work with a particular borrower is, determining the value of the credit limit, and whether the borrower will repay the debt or not. Having scored a threshold score, it can potentially receive the requested amount.

Banks use several scoring cards as they depend on credit products. For example, one card is used for cash loans, and a completely different one for car loans. According to experts, a single map can be used, but this method is inconvenient for the user. There are also several models, for different categories of persons different models are used even on the same scorecard.

The credit scoring model is a risk management tool that assesses the creditworthiness of a loan applicant by assessing its probability of default based on historical data. It uses numerical tools to rank cases using data integrated into a single value that attempts to measure risk or creditworthiness. Scoring came to the insurance industry from second-tier banks (second-tier banks), where the creditworthiness of borrowers is checked using a set of indicators. Insurance companies effectively use scoring for risk assessment, as well as in other mass protections. Unlike traditional assessment methods, scoring allows you to take into account the personal characteristics of the client - traffic police and traffic police data, information about the nature of fines, obligations, installment terms, credit history, features of investing in mobile communications, information about the circle. contacts and more.

The study of the scoring system includes:

- Systems are fully integrated, online and provide data in real time. Data capture and collections
- Data are captured digitally in real time, are fully integrated, and are managed by automated processes.
- IT/Analytics supports reporting, MIS, BI, predictive analytics, and scoring models.
- Data are used to predict behaviors, segment customers, funnel sales, assess risk, etc. Models run in real time.
- Information is used across the organization to drive business strategy and support operations, marketing and risk management.
- Products are marketed proactively and target customers according to segmented demographic, geographic, and behavioral models.

One of the first credit scoring algorithms was developed using linear programming (myFICO 2018). Initially, both the variables chosen and the scores assigned were primarily estimates. However, the systematic application of the

scoring method contributed to the uniformity of the loan application process. This approach was the beginning of the use of statistical methods for determining creditworthiness in an organized and transparent manner [9].

As with any Data Science project, when developing Scorecard, having good quality data is more important than using a better model. Creating and filtering features and replacing missing values are part of your database creation process and should be done with extreme care and attention.

The design of a credit scoring system includes: credit process analysis, system design and development, marketing, training, fine-tuning, supply and service provider.

An important advantage of the scoring system is its impartiality and the ability to standardize criteria when making a credit decision. The objectivity of the model ensures that borrowers feel that each of them is treated equally, regardless of race, gender, religion and other discriminatory factors. Moreover, this system is highly flexible, which allows the bank to use almost any credit policy that is used in the scoring algorithm.

While credit scoring ranks a borrower's credit riskiness, it does not provide an estimate of the borrower's probability of default. It simply ranks the borrower's degree of risk from highest to lowest. Thus, credit scoring suffers from the inability to determine whether Borrower A is twice as risky as Borrower B.

In general, having a good credit history can make it easier to achieve your financial and personal goals. It can be the difference between qualifying for or not being granted an important loan, such as a home mortgage or car loan. And this can directly affect how much you will have to pay in interest or fees if your approval is approved.

### **Discussion and results**

After applying for a loan, the program, having processed the data, sets a scoring score - the probability of loan approval depends on its value. Scoring systems are used not only in banks, but also in BKI (credit bureaus), insurance and leasing companies. The use of scoring systems of insurance companies leads to a reduction in the risk of loss of profitability.

The decision-making system for issuing a loan, working according to this flowchart, will receive a scoring from a certain insurance company that will effectively affect the borrower's lending, which makes it possible to assess the client's reliability.

The scoring system is the initial stage of verification. The check monitors how solvent the borrower is, whether there are errors in his credit history, whether he has taken loans before. By adding an insurance scoring system to the STB scoring system, it leads to the assessment of the borrower for insurance. If the program

refuses to receive a scoring loan, you can try to contact the bank branch again, increasing or decreasing income.

Block diagram of the work of the decision making system, when integrating bank scoring with scoring of insurance companies. When entering the client's IIN, the system refers to the following data sources.

Credit Bureau Executive Director Asem Nurgaliyeva, in her 2019 interview with Kapital.kz magazine, says: «We also use data on various restrictions from government databases. For example, the presence of a person in the lists with a travel ban, in the database of tax debtors» [10].

Making a decision only on scoring happens only in the case when the loan is very small and not for a long period. Banks check the client in several stages.

For all that, it should be noted that if a client has debts in other banks, then positive information from insurance companies will not help in approving a loan application.

In order not to be rejected for scoring, one should repay loans, utilities, pay off fines, tax debts, and pay alimony in a timely manner.

A default event in retail banking is defined as more than 90 days past due within the first 12 months after a loan is issued. Defaults are divided into fraudulent cases and ordinary defaults. By default, it is considered a fraudulent case when the delay begins in one of the first three months. This means that when applying for a loan, the borrower was not even going to return the money. Otherwise, the default is usually delinquent when the delinquency starts after the first three months of accounting.

This is why scorecards are usually divided into fraud and application scorecards. In fact, the only difference is in the definition of the target variable, while the predictor sets and data mining methods remain the same. The default cases are called «bad» cases, and the non-default cases are called «good».

We consider a dataset of 27,540 microfinance loans issued in 2012. This segment is characterized by low-money, high-margin loans with a short loan term (3–6 months). Further information about the source of the data is not provided due to a non-disclosure agreement. We develop both scorecards and test their accuracy through out-of-sample testing. The validation process requires the calculation of the performance measures (ROC curve and Gini coefficient) of the model based on a data sample that was obtained from the same parent population but was not used to develop the model itself (validation set). This approach allows the user to check the accuracy and stability of the model. The size of the test set we used was 30 % of the parental population.

Block diagram of the work of the decision making system is shown in Figure 3. Bank scoring system integrates with insurance companies scoring system.

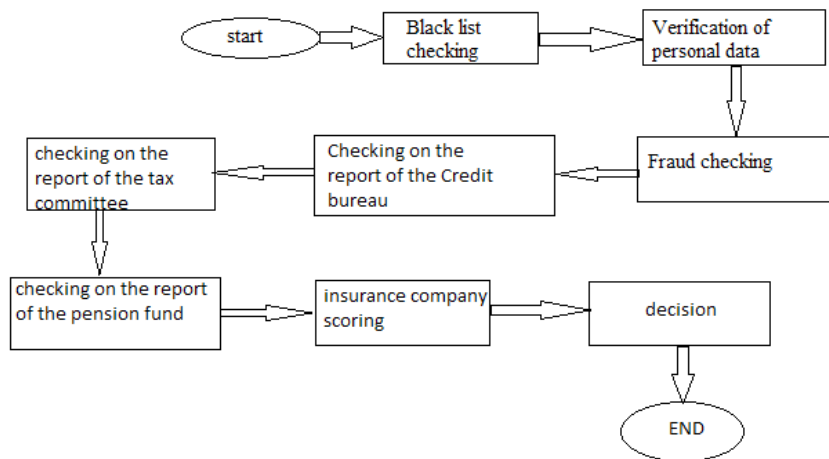


Figure 3 – Block diagram of the decision making system

When entering the client's IIN, the system refers to the following data sources:

Emergency check (black list) – available in all banks. Clients with a bad credit history who have violated lending conditions are entered into the ES array. If there is a client in an emergency, the client is denied credit.

Verification of personal data – age, gender, marital status, field of activity, real estate, etc.

Check according to the report of the State Center for the Payment of Pensions – check according to the report from the pension fund: official salary, length of service.

Checking on the report of the CB (credit bureau) – from all loans, delinquencies.

Scoring of a certain company will take place after checking the identity for fraud. And after the scoring of the insurance companies, the data is transferred to the main bank scoring. A scoring score is assigned and the decision making system displays the answer: approval or refusal.

Priority sources of information include credit history, confirmation of the borrower's income, deposits and data from the State Center for Pensions on social contributions.

The ideal high score profile looks like this; long-registered profile, a large number of subscribers, actively posts and photos, is subscribed to many groups, such as: economics and sports, visits restaurants, travels.

The information data available for building a scoring model is often referred to as a historical sample. The historical sample should reflect as accurately as

possible the population of borrowers under study, i.e. be representative. To check the adequacy and accuracy of the prediction of the scoring model at the stage of its development, the historical sample must be divided into two groups: training sample – observations, which will be directly used to build the model; test or control sample – observations from which the value of the dependent variable will be known, but they will not participate in the construction of the model, but will be used to test the accuracy of the model's prediction. The training and control samples should be formed on the basis of a random selection mechanism, usually in. Table 1 compares the criterion of good and bad customers according to the information received from social networks.

Table 1 – Comparative analysis of a positive and negative client

| options                                               | positive customer | negative customer |
|-------------------------------------------------------|-------------------|-------------------|
| Page creation date                                    | 10/18/2015        | 10/14/2022        |
| Profile                                               | open              | closed            |
| Number of subscribers                                 | over 150          | twenty            |
| Number of posts                                       | over 500 posts    | one               |
| Bad habits                                            | smokes            | I do not smoke    |
| Adrenaline junkie (motorcyclist, powder jumper, etc.) | +                 | -                 |

The test sample is used after building a logistic regression model to check its validity. For credit scoring, this is primarily the ability of the model to distinguish «good» borrowers from «bad» ones. Checking the reliability of the model consists in its application and comparison of the results on the control and test samples. The model should give correct predictions not only on the training set, but also in practice when it is applied. Typically, a two-sample model generalization strategy is used. Similar accuracy scores obtained on the training and test sets are a sign that in practice the scoring model will work in much the same way.

A more complex model generalization strategy involves the formation of three or more samples: the first sample is used to estimate the model parameters; the second sample is used to test the model, if significant deviations of the results are obtained for the training and test samples, then outliers or variables that affect the deviations are removed from them, and a new model is built using the combined first and second samples; the results of the new model are tested on the third sample.

The scoring score is a non-constant value, it can change over time. Each time the scoring system is used, the score must be recalculated and the decision

reconsidered accordingly. A threshold should be selected to determine the client's trustworthiness.

### **Conclusion**

The article discusses the main types and fundamental differences between scoring models. Advantages and method of using insurance scoring in STBs, which leads to an assessment of the borrower's reliability and potential risks of the bank. The sum of points scored by a particular client is his scoring score: if she has overcome the «bar», then the loan will be approved, if not, then the system will refuse the client. Scoring models require periodic updates, so they need to be checked and updated regularly. Correction of the settlement system in the SDS requires a mandatory adjustment when economic conditions change.

Since scoring systems usually rely on credit history, which young borrowers do not have, data from insurance companies and social networks will be able to assess the creditworthiness of the client. Based on this, the data of insurance companies and social networks increases the effectiveness of social networks.

To increase the chance of obtaining a loan, you need to emphasize your ability to pay, remove information from the page that indicates the opposite.

### **REFERENCES**

- 1 **Maria, F. V., Fernando, B.** Credit scoring in financial inclusion [Text]. – Washington : DC 20433, 2019. – 152 p.
- 2 **Anderson, S. N., Raymond, T. L.** The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation [Text]. – Oxford University Press : 2017. – 237 p.
- 3 **Fernandez, V., Dean, C., Fernando, B.** Data-Driven Segmentation in Financial Inclusion [Text]. – Washington: D. C. : CGAP. – 199 p.
- 4 **Kabakoff, S., Robert, I.** Quick R : Advanced Handling of Missing Data [Text]. – Data Camp, 2017. – 208 p.
- 5 **Schreiner, M.** Scoring : The Next Breakthrough in Microcredit [Text]. – Washington, D. C. : CGAP, 2021. – 296 p.
- 6 **Agaltsov, V. P.** Mathematical methods in programming [Text]. – Moscow : Forum, 2020. – 186 p.
- 7 **Zabotina, N. N.** Designing information systems. [Text]. – Moscow : Bustard, 2020. – 96 p.
- 8 **Chiru, C. G., Hadgu, A. T.** Sentiment-based text segmentation. [Text]. – Proc. 2nd Int. Conf. on Systems and Computer Science, 2013 – 225 p.
- 9 The World Bank Group. Credit scoring approaches guidelines. – Washington: DC 20433, 2019. – 59 p. [Electronic resource]. – <https://>

[thedocs.worldbank.org/en/doc/935891585869698451-0130022020/original/CREDITSCORINGAPPROACHESGUIDELINESFINALWEB.pdf](https://thedocs.worldbank.org/en/doc/935891585869698451-0130022020/original/CREDITSCORINGAPPROACHESGUIDELINESFINALWEB.pdf)

10 **Bondal, K.** Customer credit rating now takes into account roaming, social networks and relatives. Kapital, economic journal, 2019. [Electronic resource]. – <https://kapital.kz/finance/75898/bankovskiy-skoring-izmenilsya.html>

Material received on 06.03.23.

**\*Д. Д. Есенаман<sup>1</sup>, Л. М. Алимжанова<sup>2</sup>, А. К. Сарбасова<sup>3</sup>**

<sup>1,2,3</sup>Әл-Фараби атындағы Қазақ Ұлттық университеті,

Қазақстан Республикасы, Алматы қ.

Материал 06.03.23 баспаға түсті.

## **ТИІМДІ БАНКТІК СКОРИНГ ЖҮЙЕСІНІҢ БАҒДАРЛАМАЛЫҚ ЖАБДЫҚТАМАСЫН МОДЕЛДЕУ**

Мақала екінші деңгейлі банктердегі несиелендірудің скорингтік жүйесіне арналған. Атап айтқанда, екі скорингтік модельді біріктіру (банк және сақтандыру) қарыз алушының сенімділігін және банктің ықтимал тәуекелдерін бағалауға, сондай-ақ тәуекелділікті анықтау үшін әлеуметтік желілерді пайдалануға әкелетін шешім қабылдау жүйесі, клиенттің, Банк және сақтандыру балдары салыстырылады. Мақалада кредиттік скорингті әзірлеуде қолданылатын әдістемелер, процестер мен деректер қамтылған. Сақтандыру компанияларының деректерін пайдаланбай несиелік қабілеттілікті бағалау әдістерінің артықшылықтары сипатталған. Әлеуметтік желілерден мәліметтерді қабылдауда аналитикалық әдістер қолданылады.

Статистикалық теория нақты несиелік скорингтік жағдайларға қолданылады. Мақала қаржы секторында жұмыс істейтін адамдар үшін пайдалы. Олардың арасында несиелік қызметкерлер, тәуекел-менеджерлер мен талдаушылар, IT мамандары бар.

IT-инфрақұрылым, операциялық процедуралар, бизнес-процестер зерттелуде. Балдық жүйенің тиімділігін арттыратын балдық жүйе өнімін енгізуге арналған АТ құралдарының шолуы берілген.

Кілтті сөздер: жүйе, скоринг, модель, тәуекел, кіріс, қарыз алушы, несие беру.

**\*Д. Д. Есенаман<sup>1</sup>, Л. М. Алимжанова<sup>2</sup>, А. К. Сарбасова<sup>3</sup>**

<sup>1,2,3</sup>Казахский национальный университет имени Аль-Фараби,

Республика Казахстан, г. Алматы

Материал поступил в редакцию 06.03.23.

## **МОДЕЛИРОВАНИЕ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ЭФФЕКТИВНОЙ БАНКОВСКОЙ СКОРИНГОВОЙ СИСТЕМЫ**

*Статья посвящена скоринговой системе кредитования в банках второго уровня. А именно, система принятия решений, в которой интеграция двух скоринговых моделей (банковской и страховой) приводит к оценке надежности заемщика и потенциальных рисков банка, а также использование социальных сетей для определения рискованности клиента. Сравнивается банковский и страховой скоринг. Статья включает методологии, процессы и данные, которые используются при разработке кредитного скоринга. Описаны преимущества методик оценки кредитоспособности без использования данных страховых компаний. Используются аналитические методы, получая данные из социальных сетей.*

*Статистическая теория применяется к реальным ситуациям кредитного скоринга. Статья полезна для работающих людей, работающих в финансовой сфере. Среди них кредитные специалисты, риск-менеджеры и аналитики, IT-специалисты.*

*Изучается ИТ-инфраструктура, операционные процедуры, бизнес-процессы. Дан обзор ИТ-инструментов для реализации продукта скоринговой системы, повышающего эффективность скоринговой системы.*

*Ключевые слова: система, скоринг, модель, риск, доход, заемщик, кредитование, страхование, технология.*

Теруге 06.03.2023 ж. жіберілді. Басуға 30.03.2023 ж. қол қойылды.

Электрондық баспа

7,50 Mb RAM

Шартты баспа табағы 9,1. Таралымы 300 дана. Бағасы келісім бойынша.

Компьютерде беттеген: Е. Е. Калихан

Корректор: А. Р. Омарова, Д. А. Кожас

Тапсырыс № 4020

Сдано в набор 06.03.2023 г. Подписано в печать 30.03.2023 г.

Электронное издание

7,50 Mb RAM

Усл.печ.л. 8,4. Тираж 300 экз. Цена договорная.

Компьютерная верстка Е. Е. Калихан

Корректор: А. Р. Омарова, Д. А. Кожас

Заказ № 4020

«Toraighyrov University» баспасынан басылып шығарылған

«Торайғыров университеті» КЕ АҚ

140008, Павлодар қ., Ломов к., 64, 137 каб.

«Toraighyrov University» баспасы

«Торайғыров университеті» КЕ АҚ

140008, Павлодар қ., Ломов к., 64, 137 каб.

+7(718)267-36-69

e-mail: kereku@tou.edu.kz

www.vestnik.tou.edu.kz

<https://vestnik-pm.tou.edu.kz/>