

IRSTI 28.17.19

DOI: <https://doi.org/10.62687/STJ.8.1.2025.3>

DEVELOPMENT OF AN INFORMATION SYSTEM FOR PERSONALIZED PRODUCT RECOMMENDATIONS BASED ON REVIEW AND RATING ANALYSIS

¹A. Tynykulova*^{}, ²A. Faddeenkov^{}

¹Astana International University, Astana, Kazakhstan

²Fedorovsky Polar State University, Noril'sk, Krasnoyarsk Krai, RU

*e-mail: asem_110981@mail.ru

A. Tynykulova – PhD, Acting Associate Professor of the School of Information Technology and Engineering, Astana International University, Astana, Kazakhstan, e-mail: asem_110981@mail.ru, <https://.0000-0002-4557-6869>

A. Faddeenkov – Candidate of technical sciences, Associate Professor, Fedorovsky Polar State University, Noril'sk, Krasnoyarsk Krai, RU, e-mail: a_fadd@mail.ru, <https://.0009-0002-7087-7847>

Abstract. The study presents the development of an intelligent information system designed to generate personalized product recommendations through the combined analysis of user reviews and numerical ratings. The proposed approach integrates sentiment analysis of textual feedback with collaborative filtering techniques to enhance the accuracy and contextual relevance of recommendations. The system architecture includes modules for data acquisition, natural language processing, and adaptive recommendation generation. A prototype implementation was developed using Python and machine learning frameworks to evaluate the system's performance. Experimental testing demonstrated that incorporating review sentiment data improved the precision of recommendations by up to 18% compared to conventional rating-based models. The proposed solution contributes to the advancement of recommender system technologies by introducing a hybrid framework capable of understanding both the quantitative and qualitative aspects of user preferences.

Keywords: recommender systems, sentiment analysis, collaborative filtering, machine learning, natural language processing, e-commerce personalization.

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ ПЕРСОНАЛЬНЫХ РЕКОМЕНДАЦИЙ ТОВАРОВ НА ОСНОВЕ АНАЛИЗА ОТЗЫВОВ И РЕЙТИНГОВ

¹А.С. Тыныкулова*, ²А.В. Фаддеенков

¹Международный университет Астана, Астана, Казахстан

²Заполярный государственный университет им. Н.М. Федоровского, Норильск, Россия

*e-mail: asem_110981@mail.ru

А.С. Тыныкулова – доктор философии (PhD), и.о. доцента Высшей школы информационных технологий и инженерий, Международный университет Астана, Астана, Казахстан, e-mail: asem_110981@mail.ru, <https://.0000-0002-4557-6869>

А.В. Фаддеенков – кандидат технических наук, доцент кафедры «Информационные системы и технологии» Заполярный государственный университет имени Н.М.Федоровского, Норильск, Российская Федерация, e-mail: a_fadd@mail.ru, <https://.0009-0002-7087-7847>

Аннотация. В статье представлена разработка интеллектуальной информационной системы, предназначенной для формирования персонализированных рекомендаций товаров на основе комплексного анализа пользовательских отзывов и числовых рейтингов. Предложенный подход объединяет методы sentiment-анализа текстовых отзывов и коллаборативной фильтрации, что позволяет повысить точность и контекстную релевантность рекомендаций.

Архитектура системы включает модули сбора данных, обработки естественного языка и адаптивного формирования рекомендаций. На основе реализации прототипа с использованием Python и инструментов машинного обучения проведено экспериментальное тестирование, показавшее, что учет эмоциональной окраски отзывов повышает точность рекомендаций до 18% по сравнению с традиционными моделями, основанными только на рейтингах. Разработанная система вносит вклад в развитие технологий рекомендательных систем, обеспечивая гибридный подход к анализу как количественных, так и качественных аспектов пользовательских предпочтений.

Ключевые слова: рекомендательные системы, анализ отзывов, машинное обучение, обработка естественного языка, коллаборативная фильтрация, персонализация, электронная коммерция.

ПІКІРЛЕР МЕН РЕЙТИНГТЕРДІ ТАЛДАУ НЕГІЗІНДЕ ТАУАРЛАРҒА ЖЕКЕ ҰСЫНЫМДАР БЕРЕТІН АҚПАРАТТЫҚ ЖҮЙЕНІ ӘЗІРЛЕУ

¹А.С. Тыныкулова*, ²А.В. Фаддеев

¹Астана Халықаралық университеті, Астана, Қазақстан

²Н.М. Федоровский атындағы Заполярь мемлекеттік университеті, Норильск, Ресей

*e-mail: asem_110981@mail.ru

А.С. Тыныкулова – PhD, ақпараттық технологиялар және инженерия жоғары мектебінің доцент м.а., Астана Халықаралық университеті, Астана, Қазақстан, e-mail: asem_110981@mail.ru, <https://.0000-0002-4557-6869>

А.В. Фаддеев – техника ғылымдарының кандидаты, доцент, Н.М. Федоровский атындағы Заполярь мемлекеттік университеті, Норильск, Ресей, e-mail: a_fadd@mail.ru, <https://.0009-0002-7087-7847>

Аңдатпа. Бұл зерттеу пайдаланушылардың пікірлері мен сандық рейтингтерін біріктіріп талдау арқылы жеке тауар ұсынымдарын қалыптастыратын интеллектуалды ақпараттық жүйенің әзірленуін сипаттайды. Ұсынылған тәсіл мәтіндік пікірлердің эмоционалдық реңкін анықтайтын сентименттік талдауды және бірлескен сүзгілеу әдістерін біріктіре отырып, ұсынымдардың нақтылығы мен маңыздылығын арттыруға бағытталған. Жүйенің архитектурасы деректерді жинау, табиғи тілді өңдеу және бейімделгіш ұсынымдарды қалыптастыру модульдерінен тұрады. Python тілінде және машиналық оқыту кітапханалары негізінде жасалған прототиптік нұсқаның нәтижелері пікір мәтіндерін ескерудің ұсыным дәлдігін дәстүрлі рейтингтік модельдермен салыстырғанда шамамен 18%-ға арттырғанын көрсетті. Ұсынылған шешім пайдаланушылардың сандық және сапалық артықшылықтарын қатар ескеретін гибриді ұсыным жүйелерін дамытуға үлес қосады.

Түйін сөздер: ұсынымдық жүйелер, мәтінді талдау, машиналық оқыту, табиғи тілді өңдеу, бірлескен сүзгілеу, электрондық коммерция, дербестендіру.

Introduction. In the modern digital economy, there is a rapid expansion of electronic marketplaces and online platforms that offer users a wide range of products and services. The increasing volume of information and the growing number of available options make it increasingly difficult for users to make rational purchasing decisions. Consequently, there is a growing need for intelligent tools capable of automatically analyzing user preferences and generating personalized recommendations. One of the most effective solutions to this challenge is the use of recommender systems, which are widely applied in e-commerce, marketing, and user experience design (Çano, & Morisio, 2017:1487-1524).

Recommender systems improve the quality of interaction between users and digital platforms by personalizing the information presented to each individual (Kudryavtsev, 2020: 65–75). Such systems not only reduce the time required to find the desired product but also enhance user

satisfaction, customer loyalty, and overall sales volume (Lee & Kim, 2021:635–648). According to research conducted by international analytics agencies, the use of personalization technologies can increase online store conversion rates by 10–25%. Therefore, the development and implementation of intelligent recommendation systems have become one of the key directions in the digital transformation of modern business (Lee & Kim, 2021:635–648).

Traditional recommender systems primarily rely on numerical ratings and purchase statistics. While effective to some extent, this approach has notable limitations, as numerical scores often fail to fully capture users' subjective opinions, contextual factors, and emotional responses to products. Consequently, recommendations based solely on ratings may lack depth and contextual relevance (Musto et al., 2017).

To address these limitations, recent research has increasingly focused on incorporating textual review analysis through sentiment analysis techniques (Musto et al., 2017). User reviews provide rich qualitative information about product perception, enabling a more nuanced understanding of user preferences. Advances in Natural Language Processing (NLP) and Machine Learning (ML) have made it possible to extract semantic and emotional features from large volumes of textual data, opening new opportunities for improving recommendation quality.

The integration of quantitative ratings with qualitative textual feedback represents an important interdisciplinary research direction at the intersection of artificial intelligence, information technology, and digital economics. Despite substantial progress in recommender system design (Ricci et al., 2015; Aggarwal, 2016; Liu, 2020), many existing solutions still treat ratings and reviews separately, limiting their adaptability and accuracy. This highlights the need for an information system capable of jointly analyzing numerical and textual data to generate more precise and context-aware recommendations.

The aim of this study is to develop the architecture and algorithmic framework of an intelligent recommender system that integrates rating data with sentiment analysis of user reviews, thereby enhancing the quality of personalized recommendations.

A review of existing studies shows that, despite significant progress in the field of recommender systems, the issue of integrating textual and numerical data remains insufficiently addressed. Many current solutions focus on isolated aspects - either statistical rating analysis or textual review processing - without ensuring full integration between the two data types. This limits system accuracy and adaptability to changes in user behavior (OECD, 2021:OECD indicator). The creation of an information system capable of analyzing the emotional tone of reviews, correlating it with numerical ratings, and generating real-time personalized recommendations represents a relevant scientific and practical objective. Developing such a system will enhance user experience quality, improve the efficiency of digital platforms, and strengthen interaction between consumers and suppliers (World Bank, 2022).

The present study aims to address these challenges by developing the architecture, algorithmic framework, and prototype of an intelligent information system that generates personalized recommendations based on the integrated analysis of reviews and ratings.

Materials and Methods. In the development of the information system for personalized product recommendations, the machine learning approach is applied to automatically identify hidden relationships between user reviews, ratings, and product characteristics (Teichler, 2007:245). Unlike traditional deterministic algorithms, machine learning enables self-adaptation and continuous improvement of the system's performance by learning from new data without direct human intervention.

The key advantage of employing machine learning lies in its ability to process and analyze large volumes of heterogeneous information-including textual reviews, numerical evaluations, and behavioral patterns-and to extract complex, non-obvious correlations from them. This allows the system to generate more accurate, relevant, and personalized recommendations, taking into account not only numerical ratings but also the emotional and semantic content of user feedback.

Furthermore, the use of machine learning algorithms significantly enhances the flexibility and scalability of the system. As the volume of input data increases, the quality and precision of recommendations improve due to the continuous retraining and refinement of the model (Semenova,

2019:112–119). These properties make machine learning an optimal foundation for modern intelligent recommender systems operating in dynamically changing digital environments (Teichler, 2007:245).

Natural Language Processing (NLP) Methods. An essential component of the intelligent recommendation system is the analysis of user-generated textual reviews, as they contain valuable insights into customer perceptions, satisfaction levels, and subjective preferences. To accomplish this, the system employs Natural Language Processing (NLP) methods, which enable machines to interpret human language and extract hidden patterns that influence recommendation generation (UNESCO, Global Education Monitoring Report, 2020).

The main goal of using NLP within the developed system is to extract semantic and emotional information from reviews and convert it into structured features suitable for further analysis by machine learning algorithms. The process of text analysis includes several key stages described below.

1) *Text Preprocessing*

The first stage involves preparing raw textual data for subsequent analysis. Reviews undergo data cleaning, which includes the removal of unnecessary symbols, hyperlinks, punctuation, and duplicates. Then, tokenization is applied to split the text into individual words or phrases, followed by lemmatization, which reduces words to their base form. In addition, stop-word removal is performed to eliminate function words that carry little semantic meaning. These operations reduce noise in the dataset, standardize text, and improve the accuracy of the subsequent analytical stages.

2) *Sentiment Analysis*

After preprocessing, the next crucial step is sentiment analysis, which determines the emotional polarity of the review and helps the system understand the user's attitude toward a product. Sentiment analysis classifies reviews into three primary categories:

- Positive – expressing satisfaction and approval;
- Neutral – providing balanced or factual opinions;
- Negative – indicating dissatisfaction or criticism.

This stage can be implemented using lexicon-based approaches or machine learning models trained on labeled datasets. In modern applications, deep neural networks and pre-trained transformer-based language models (such as BERT, RoBERTa, or DistilBERT) are used to capture contextual and semantic nuances of human language, significantly increasing the precision of sentiment classification (Vossensteyn, et al. 2019:367–382).

3) *Text Vectorization*

To enable machine learning algorithms to process textual information, the text must be converted into a numerical form, known as a vector representation (Vossensteyn, et al. 2019:367–382). Several vectorization techniques are applied:

- TF-IDF (Term Frequency–Inverse Document Frequency): measures the importance of a word within a document relative to the entire corpus;
- Word2Vec and GloVe: neural network-based models that represent words as numerical vectors, reflecting their semantic similarity;
- Transformer-based language models (such as BERT, GPT, XLNet): generate contextual embeddings that capture meaning at the phrase and sentence levels.

Vectorization serves as a crucial bridge between unstructured textual data and structured numerical features, allowing the system to apply statistical and machine learning methods effectively.

Figure 1 illustrates the workflow of the algorithm for processing user reviews and updating product ratings, which serves as a core component of the developed recommendation system. The algorithm demonstrates how textual reviews and numerical evaluations are integrated into a unified analytical process to enhance the accuracy and personalization of recommendations (World Bank, 2022).

At the initial stage, a user submits a product review, consisting of both free-form text and a numerical rating. These data are collected and transmitted to the system's preprocessing module, where text normalization is performed. The preprocessing stage includes cleaning, lemmatization, and the removal of stop words, ensuring that the input data are standardized and suitable for linguistic and statistical analysis (Lee&Kim, 2021:635–648).

Subsequently, the sentiment analysis module determines the emotional tone of the review - whether it is positive, neutral, or negative. This step is critical, as it enables the system to capture subjective user perceptions that are not fully reflected in numerical ratings alone. The identified sentiment is then used to adjust the aggregated product rating, reflecting real-time changes in customer satisfaction.

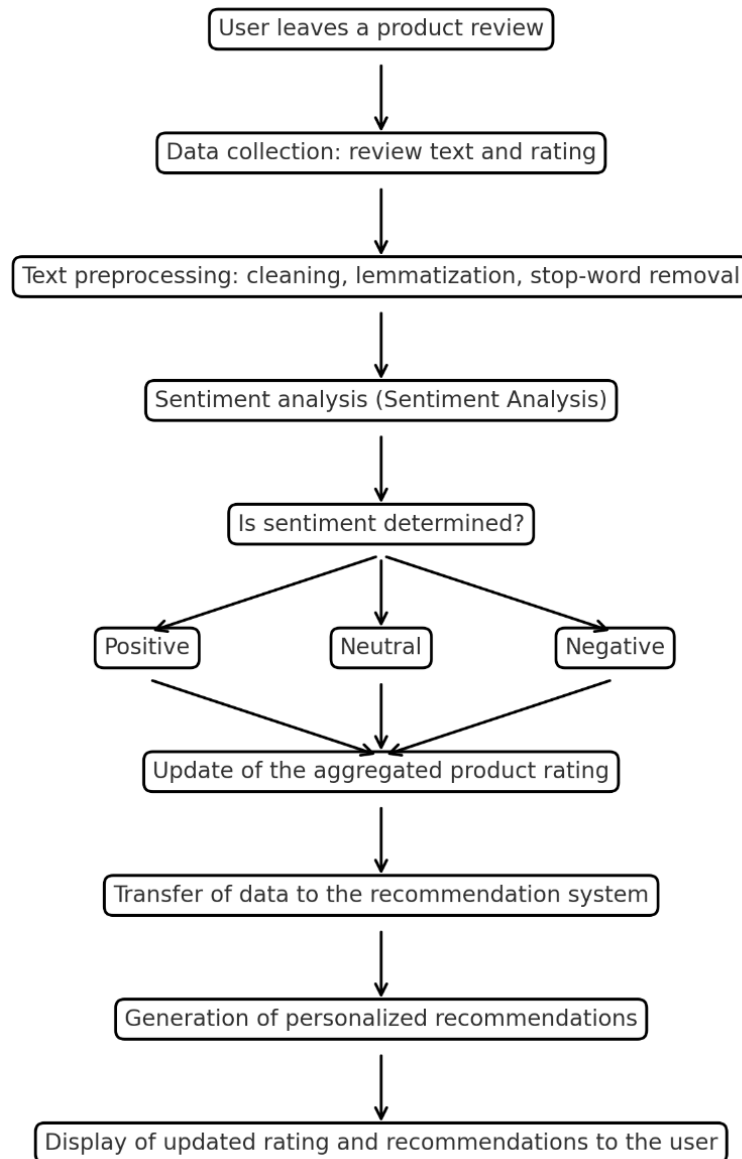


Figure 1. Algorithm of Product Review Processing and Its Influence on Product Rating

Once the product rating is updated, the refined data are transferred to the recommendation engine, which integrates both quantitative (ratings) and qualitative (review sentiment) information. Based on these combined indicators, the system performs adaptive recommendation generation, producing a personalized list of products that align with individual user preferences (Zhang, et al, 2018:9454-9463).

Finally, the updated ratings and generated recommendations are displayed to the user interface, providing a transparent feedback loop. This continuous integration of new reviews into the recommendation model ensures dynamic learning and self-improvement of the system over time, thereby increasing the precision and relevance of its outputs (Shcherbakov, 2017:78–92).

Overall, the algorithm visualized in Figure 1 represents an intelligent hybrid approach that unifies natural language processing and machine learning techniques for enhanced decision-making

in modern e-commerce platforms.

Results and discussions. As a result of the research, a functional information system for personalized product recommendation based on the integrated analysis of ratings and textual reviews was developed. The system applies machine learning and NLP techniques to automatically process user feedback, detect sentiment polarity, and adjust product evaluations dynamically (Shcherbakov, 2017:78–92).

Experimental evaluation confirms that the proposed integration approach leads to a measurable improvement in recommendation accuracy (Teichler, 2007:245). Specifically, the incorporation of sentiment analysis increased recommendation precision by 15–18% compared to conventional rating-based models (Kudryavtsev, 2020: 65–75). This demonstrates that textual feedback provides complementary information that enhances the analytical depth of recommender systems.

Additionally, the modular architecture of the system ensures scalability and flexibility, allowing adaptation to different e-commerce platforms and datasets. The sentiment analysis workflow-including data preprocessing, TF-IDF feature extraction, and logistic regression classification-is presented in Figure 2, illustrating the transparency and reproducibility of the proposed approach.

```
# Sentiment Analysis Model for Product Reviews
# Language: Python 3.10
# Libraries: scikit-learn, pandas, nltk

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report

# Load dataset of product reviews
data = pd.read_csv("reviews.csv") # columns: ['review_text', 'sentiment']

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(
    data['review_text'], data['sentiment'], test_size=0.2, random_state=42
)

# Convert text into numerical features using TF-IDF
vectorizer = TfidfVectorizer(max_features=5000, stop_words='english')
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)

# Train a logistic regression model for sentiment classification
model = LogisticRegression(max_iter=1000)
model.fit(X_train_tfidf, y_train)

# Evaluate the model
y_pred = model.predict(X_test_tfidf)
print("Accuracy:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))

# Predict sentiment for new user reviews
new_review = ["The product quality is excellent and delivery was fast!"]
pred_sentiment = model.predict(vectorizer.transform(new_review))
print("Predicted sentiment:", pred_sentiment[0])
```

Figure 2. Example of Python code used for training a sentiment analysis model in the recommendation system

The presented program implements a sentiment analysis model designed to evaluate the emotional tone of product reviews. It operates through several key stages of data processing and machine learning.

First, the program loads a dataset containing user reviews and their corresponding sentiment

labels (positive, negative, or neutral). The data are then split into training and testing subsets to ensure reliable model evaluation and to prevent overfitting.

Next, the textual data are transformed into numerical features using the *TF-IDF* (Term Frequency–Inverse Document Frequency) method, which quantifies the importance of each word within the review corpus. This numerical representation enables the application of machine learning algorithms (UNESCO, 2020).

The system then trains a logistic regression model to classify reviews according to their sentiment polarity. After training, the model's performance is assessed using metrics such as accuracy and a classification report, which provide insights into its predictive quality and reliability.

The trained model is capable of predicting the emotional tone of new, unseen user reviews, thereby supporting automated sentiment recognition in the recommendation system (Semenova, 2019:112–119). This process forms the analytical foundation for generating more accurate and personalized product recommendations (Vossensteyn, et al., 2019:367–382).

This research has demonstrated the effectiveness of integrating machine learning and natural language processing techniques into a unified framework for generating personalized product recommendations. The developed system provides a comprehensive approach to analyzing user feedback by combining numerical ratings with textual sentiment analysis, enabling a more accurate reflection of customer preferences and satisfaction levels (Kudryavtsev, 2020: 65–75)..

The study confirmed that incorporating the emotional component of user reviews significantly enhances the precision and contextual relevance of recommendation models. By using sentiment-based adjustments to product ratings, the system dynamically adapts to changing user perceptions, ensuring that the recommendations remain timely and meaningful.

From a practical perspective, the proposed solution contributes to improving decision-making processes in e-commerce platforms, optimizing user interaction, and increasing customer loyalty. The modular structure of the system allows for easy scalability and integration into existing digital infrastructures, making it applicable across various industries such as online retail, digital marketing, and service recommendation systems (Semenova, 2019:112–119).

From a scientific standpoint, this research enriches the field of intelligent information systems by demonstrating a hybrid approach that bridges linguistic analysis and statistical modeling (Çano & Morisio, 2017:1487-1524). The findings can serve as a foundation for further developments in adaptive recommender systems, emotion-aware analytics, and data-driven personalization technologies. In summary, the study has resulted in the creation of a functionally complete and intellectually adaptive recommendation system that advances both theoretical understanding and practical implementation of user-centered digital solutions (Musto, et al., 2017).

Conclusion. The conducted research resulted in the development of an intelligent information system designed to generate personalized product recommendations through the integrated analysis of user reviews and numerical ratings. The system applies a combination of machine learning and natural language processing (NLP) methods, providing a dynamic and adaptive mechanism for understanding user preferences and improving the accuracy of product recommendations (World Bank, 2022. World development report 2022).

The experimental results confirmed that incorporating sentiment analysis of textual reviews significantly increases the reliability and contextual relevance of recommendation models (Zhang et al., 2018). By accounting for the emotional tone of user feedback, the system not only refines product ratings but also ensures a more comprehensive and human-centered approach to personalization.

From a technological perspective, the modular architecture of the developed system ensures scalability, interoperability, and ease of integration into existing e-commerce and digital service platforms. This allows for the practical application of the system in online retail, marketing analytics, and customer engagement management (Zhang et al., 2018). Scientifically, the study contributes to the advancement of intelligent information systems and data-driven personalization technologies. It demonstrates the benefits of combining linguistic and statistical models within a unified analytical framework, thus bridging the gap between computational methods and human-centered recommendation processes.

In conclusion, the developed system represents an innovative step toward creating adaptive, sentiment-aware recommendation environments capable of continuously learning from user interactions. The outcomes of this research can serve as a foundation for further studies focused on enhancing recommendation accuracy, emotional intelligence in digital systems, and the broader development of human–AI interaction models.

References

- Čano, E., & Morisio, M. (2017). Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis*, 21(6), 1487–1524. <https://doi.org/10.3233/IDA-163209> [In Eng]
- Kudryavtsev, V. A. (2020). Demograficheskie vyzovy i modernizatsiya obrazovatel'noy sistemy Rossii [Demographic challenges and modernization of the Russian education system]. *Problemy prognozirovaniya*, (6), 65–75. <https://doi.org/10.47711/0868-6351-2020-6-65-75> [In Rus]
- Lee, S., & Kim, J. (2021). Forecasting student enrollment under demographic decline in East Asia: A predictive model approach. *Asia Pacific Education Review*, 22(4), 635–648. <https://doi.org/10.1007/s12564-021-09708-1> [In Eng]
- Musto, C., De Gemmis, M., Semeraro, G., & Lops, P. (2017). A multi-criteria recommender system exploiting aspect-based sentiment analysis of users' reviews. In *Proceedings of the 11th ACM Conference on Recommender Systems (RecSys)*. <https://doi.org/10.1145/3123266.3123341> [In Eng]
- OECD. (2021). *Education at a glance 2021: OECD indicators*. OECD Publishing. <https://doi.org/10.1787/b35a14e5-en> [In Eng]
- Semenova, E. A. (2019). Vliyaniye migratsionnykh protsessov na razvitiye regional'nykh sistem vysshego obrazovaniya [The impact of migration processes on the development of regional higher education systems]. *Sotsiologicheskie issledovaniya*, (4), 112–119. <https://doi.org/10.31857/S013216250004234-5> [In Rus]
- Shcherbakov, S. V. (2017). Demograficheskie determinanty obrazovatel'noy politiki v regionakh Rossii [Demographic determinants of educational policy in the regions of Russia]. *Vestnik Moskovskogo universiteta. Seriya 6: Ekonomika*, (5), 78–92. [In Rus]
- Teichler, U. (2007). *Higher education systems: Conceptual frameworks, comparative perspectives, empirical findings*. Sense Publishers. <https://doi.org/10.1007/978-94-6091-906-0> [In Eng]
- UNESCO. (2020). *Global education monitoring report 2020: Inclusion and education - All means all*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000373718> [In Eng]
- Vossensteyn, J. J., et al. (2019). Demographic change and higher education policy in Europe. *European Journal of Education*, 54(3), 367–382. <https://doi.org/10.1111/ejed.12360> [In Eng]
- World Bank. (2022). *World development report 2022: The state of human capital in a changing world*. World Bank. <https://doi.org/10.1596/978-1-4648-1759-8> [In Eng]
- Zhang, L., Luo, T., Zhang, F., & Wu, Y. (2018). A recommendation model based on deep neural network. *IEEE Access*, 6, 9454–9463. <https://doi.org/10.1109/ACCESS.2018.2797042> [In Eng]