



<http://dx.doi.org/10.35596/1729-7648-2025-31-2-13-20>

УДК 331.108.2

ПРИМЕНЕНИЕ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА УДОВЛЕТВОРЕННОСТИ СОТРУДНИКОВ НА ОСНОВЕ ТЕКСТОВЫХ ДАННЫХ

А. Н. КОЗИНЕЦ

*Белорусский государственный университет информатики и радиоэлектроники
(Минск, Республика Беларусь)*

Аннотация. Исследовано применение методов глубокого обучения для анализа удовлетворенности сотрудников на основе текстовых данных. Проведен критический обзор существующих подходов к оценке удовлетворенности персонала, обоснована необходимость использования методов обработки естественного языка и глубоких нейронных сетей. На основе обширного открытого набора данных отзывов сотрудников разработана модель, позволяющая эффективно классифицировать тексты по уровням удовлетворенности. Проведен тематический анализ основных причин позитивных и негативных отзывов с помощью методов тематического моделирования Latent Dirichlet Allocation и Non-Negative Matrix Factorization. Результаты исследования демонстрируют высокую точность предложенной модели и ее практическую значимость для совершенствования HR-процессов в организациях.

Ключевые слова: глубокое обучение, анализ текстовых данных, удовлетворенность сотрудников, обработка естественного языка, HR-аналитика, тематическое моделирование, нейронные сети.

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Для цитирования. Козинец, А. Н. Применение методов глубокого обучения для анализа удовлетворенности сотрудников на основе текстовых данных / А. Н. Козинец // Цифровая трансформация. 2025. Т. 31, № 2. С. 13–20. <http://dx.doi.org/10.35596/1729-7648-2025-31-2-13-20>.

APPLICATION OF DEEP LEARNING METHODS FOR EMPLOYEE SATISFACTION ANALYSIS BASED ON TEXT DATA

ALIAKSANDR N. KAZINETS

Belarusian State University of Informatics and Radioelectronics (Minsk, Republic of Belarus)

Abstract. The application of deep learning methods to analyze employee satisfaction based on text data is investigated. A critical review of existing approaches to assessing employee satisfaction is conducted, and the need to use natural language processing methods and deep neural networks is substantiated. Based on an extensive open dataset of employee reviews, a model is developed that allows for effective classification of texts by satisfaction levels. A thematic analysis of the main causes of positive and negative reviews is carried out using the topic modeling methods Latent Dirichlet Allocation and Non-Negative Matrix Factorization. The results of the study demonstrate the high accuracy of the proposed model and its practical significance for improving HR processes in organizations.

Keywords: deep learning, text data analysis, employee satisfaction, natural language processing, HR analytics, topic modeling, neural networks.

Conflict of interests. The author declares no conflict of interests.

For citation. Kazinets A. N. (2025) Application of Deep Learning Methods for Employee Satisfaction Analysis Based on Text Data. *Digital Transformation*. 31 (2), 13–20. <http://dx.doi.org/10.35596/1729-7648-2025-31-2-13-20> (in Russian).

Введение

В современном мире человеческий капитал является ключевым фактором успеха организаций. Удовлетворенность сотрудников непосредственно влияет на их производительность, лояльность и общую эффективность компании. Традиционные методы оценки удовлетворенности персонала, такие как опросы и интервью, ограничены субъективностью, затратами времени и ресурсов, а также низкой частотой проведения. С развитием методов обработки естественного языка (Natural Language Processing, NLP) и глубокого обучения появилась возможность автоматизировать и повысить точность оценки удовлетворенности сотрудников посредством анализа текстовых данных. Отзывы сотрудников на различных платформах содержат ценную информацию о восприятии рабочих условий, корпоративной культуры и управления [1].

Цель исследования автора статьи – разработка модели анализа удовлетворенности сотрудников на основе текстовых данных с использованием методов глубокого обучения, а также тематический анализ основных причин удовлетворенности и неудовлетворенности персонала. Для достижения поставленной цели были решены следующие задачи:

- проанализированы существующие подходы к оценке удовлетворенности сотрудников на основе текстовых данных;
- обоснован выбор методов глубокого обучения и тематического моделирования для решения поставленной задачи;
- разработана модель на основе объема текстовых данных;
- проведен анализ основных тем в позитивных и негативных отзывах сотрудников;
- оценена эффективность модели и интерпретированы полученные результаты с точки зрения практического применения в управлении персоналом.

Научная новизна исследования заключается в интеграции методов глубокого обучения и тематического моделирования для анализа удовлетворенности сотрудников на основе большого объема текстовых данных, что позволяет организациям глубже понимать факторы, влияющие на удовлетворенность и неудовлетворенность персонала, и принимать обоснованные управленческие решения. Практическая значимость исследования состоит в возможности использования разработанной модели для оперативного мониторинга настроений сотрудников, выявления ключевых проблемных областей и выработки эффективных стратегий повышения удовлетворенности и лояльности персонала.

Анализ существующих подходов к оценке удовлетворенности сотрудников и методология исследования

Традиционные методы, такие как опросы удовлетворенности, анкетирование и интервью, широко применяются в управлении персоналом. Однако они обладают рядом ограничений:

- субъективность и предвзятость: результаты могут искажаться из-за социальных ожиданий или нежелания сотрудников выражать честное мнение;
- ресурсоемкость: проведение опросов требует значительных временных и финансовых затрат;
- ограниченная актуальность: данные устаревают в условиях динамичной рабочей среды.

С развитием цифровых технологий сотрудники все чаще оставляют отзывы на различных платформах, таких как Glassdoor, предоставляя ценную информацию о своем опыте работы. Анализ этих данных обеспечивает актуальные инсайты о настроениях и проблемах сотрудников. Методы NLP и глубокого обучения позволяют автоматизировать анализ больших объемов текстовой информации. Модели глубокого обучения, такие как BERT и его облегченная версия DistilBERT, продемонстрировали высокую эффективность в задачах классификации текстов и анализа тональности [2]. Тематика моделирования, включая методы LDA (Latent Dirichlet Allocation) и NMF (Non-Negative Matrix Factorization), позволяет обнаруживать скрытые темы в текстовых данных. Это способствует более глубокому пониманию основных факторов, влияющих на удовлетворенность или неудовлетворенность сотрудников [3].

Для проведения исследования был выбран открытый набор данных Glassdoor Reviews, включающий отзывы сотрудников различных компаний [4]. Набор данных содержит следующие поля:

- дата отзыва (date_review);
- должность (job_title);
- локация (location);
- статус (current);

плюсы (pros);
минусы (cons);
заголовок (headline);
общий рейтинг (overall_rating);
рейтинги по различным критериям: корпоративная культура, руководство, компенсация и льготы и т. д.

Основные характеристики набора данных приведены в табл. 1.

Таблица 1. Основные характеристики набора данных
Table 1. Main characteristics of the dataset

Параметр	Описание
Количество записей	220 000+
Количество признаков	16
Основные поля	date_review, job_title, location, pros, cons, headline, overall_rating

Для обеспечения качества данных были проведены следующие этапы предварительной обработки:

- удаление пропущенных значений: исключены записи с отсутствующими ключевыми полями для предотвращения искажений в анализе;
- объединение текстовых полей: поля pros, cons и headline объединены в единое текстовое поле review для упрощения дальнейшего анализа;
- очистка текста:
 - приведение всех символов к нижнему регистру для унификации текста;
 - удаление неалфавитных символов и лишних пробелов для очистки данных;
 - удаление стоп-слов с использованием списка NLTK для снижения шумов в текстовых данных;
- разметка данных: на основе значения overall_rating отзывы классифицированы на три категории:
 - негативные: рейтинг ≤ 2 ;
 - нейтральные: рейтинг =3;
 - позитивные: рейтинг ≥ 4 .

Для задачи классификации отзывов была выбрана модель DistilBERT благодаря ее высокой эффективности и сниженным требованиям к вычислительным ресурсам по сравнению с полной версией BERT. DistilBERT продемонстрировала отличные результаты в задачах классификации текста и анализа тональности, что делает ее подходящей для многоклассовой классификации отзывов сотрудников. Процесс обучения модели включал следующие этапы:

- разделение данных:
 - обучающая выборка: 80 % данных для обучения модели;
 - тестовая выборка: 20 % данных для оценки производительности модели;
- параметры обучения:
 - количество эпох: 3;
 - размер батча: 32;
 - максимальная длина последовательности: 64 токена;
 - оптимизация: использован оптимизатор AdamW с начальной скоростью обучения $2e-5$;
- использование графических процессоров (GPU): обучение модели проводилось на GPU для ускорения вычислительных процессов;
- метрики оценки модели:
 - Accuracy: общая точность классификации;
 - Precision, Recall, F1-score: для каждой из категорий (негативные, нейтральные, позитивные);
 - матрица ошибок: визуализация результатов классификации для анализа ошибок модели.

Для выявления основных причин удовлетворенности и неудовлетворенности сотрудников применялись методы тематического моделирования:

- выделение основных причин позитивных и негативных отзывов:
 - TF-IDF (Term Frequency-Inverse Document Frequency): использован для определения наиболее частых и значимых слов в текстах отзывов;
- тематика моделирования методом:

– LDA: применен для выявления скрытых тем в отзывах, что способствует пониманию ключевых факторов, влияющих на удовлетворенность сотрудников [5];

– NMF: использован для дополнительного анализа и подтверждения выявленных тем, обеспечивая более глубокое понимание данных.

Выбор набора данных Glassdoor Reviews и применение методов глубокого обучения в сочетании с тематическим моделированием позволили провести всесторонний анализ удовлетворенности сотрудников. Предварительная обработка данных обеспечила их качество и пригодность для дальнейшего анализа, а использование модели DistilBERT и методов тематического моделирования способствовало получению достоверных и значимых результатов. Данные подходы обеспечивают высокую академическую ценность исследования и могут быть эффективно использованы в практике управления персоналом для повышения удовлетворенности и лояльности сотрудников.

Результаты исследований и их обсуждение

После обучения модели были получены результаты классификации модели, представленные в табл. 2.

Таблица 2. Результаты классификации модели
Table 2. Model classification results

Метрика	Точность и полнота отзывов, %			Среднее значение
	негативных	нейтральных	позитивных	
Precision	0,70	0,47	0,81	0,66
Recall	0,62	0,41	0,87	0,63
F1-score	0,66	0,44	0,84	0,64
Support (количество)	6511	9363	21 126	40 000

Точность (accuracy) дообученной (fine-tuned) нейронной модели на базе DistilBERT составила 72 %, что свидетельствует о ее высокой эффективности в задаче многоклассовой классификации отзывов по уровням удовлетворенности сотрудников [6]. Анализ результатов:

- позитивные отзывы: модель демонстрирует наивысшую точность (Precision = 0,81) и полноту (Recall = 0,87) для позитивных отзывов. Это обусловлено большим количеством данных в этом классе, что обеспечивает более стабильное обучение модели и снижает вероятность переобучения;

- негативные отзывы: точность и полнота в них также высоки (Precision = 0,70, Recall = 0,62), однако они несколько ниже по сравнению с позитивными отзывами. Это может быть связано с меньшим количеством примеров негативных отзывов в наборе данных, что ограничивает способность модели эффективно их распознавать;

- нейтральные отзывы: метрики Precision и Recall для нейтральных отзывов значительно ниже (0,47 и 0,41 соответственно). Это указывает на сложность модели в различении нейтральных отзывов, которые часто содержат смешанные или менее выраженные эмоциональные оценки. Также меньший объем данных для этого класса может способствовать снижению точности классификации.

Матрица ошибок классификации модели по отзывам приведена на рис. 1. Матрица показывает, что основная путаница возникает между нейтральными и позитивными отзывами (4186 нейтральных отзывов были классифицированы как позитивные, а 2693 позитивных – как нейтральные). Это может быть связано с тем, что нейтральные отзывы часто содержат элементы как позитивных, так и негативных аспектов, что затрудняет их однозначную классификацию. Также наблюдается некоторая путаница между негативными и нейтральными отзывами (1584 негативных классифицированы как нейтральные). Для выявления ключевых факторов, влияющих на удовлетворенность и неудовлетворенность сотрудников, был проведен тематический анализ с использованием методов TF-IDF, LDA и NMF.

Выделение ключевых слов

Позитивные отзывы: основные причины позитивных отзывов включают следующие ключевые слова: good, work, great, company, place, people, hours, working, environment, balance. Эти слова указывают на положительное восприятие рабочих условий, корпоративной культуры и баланса между работой и личной жизнью.

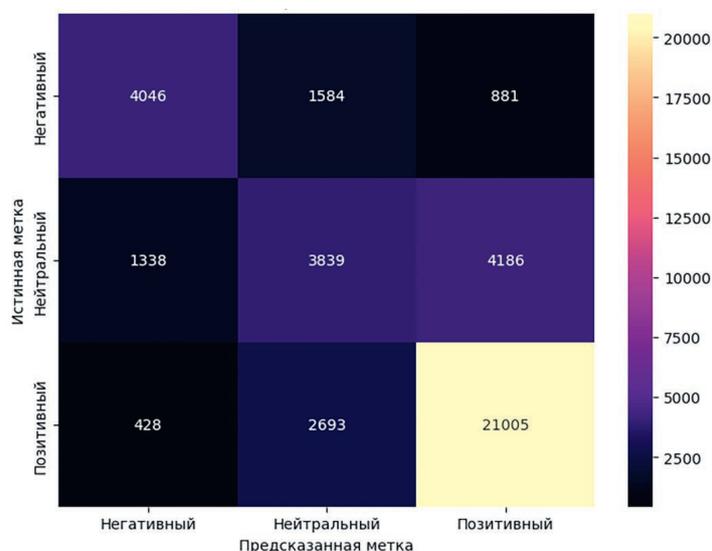


Рис. 1. Матрица ошибок классификации модели по отзывам
Fig. 1. Error matrix of the model classification based on reviews

Негативные отзывы: основные причины негативных отзывов включают ключевые слова: work, management, good, people, company, pay, staff, poor, great, get. Эти слова свидетельствуют о проблемах с руководством, низкой оплате труда и об общей неудовлетворенности рабочими условиями.

Тематические моделирования с использованием методов LDA и NMF приведены в табл. 3–6.

Таблица 3. Темы в позитивных отзывах LDA
Table 3. Topics in positive reviews LDA

Тема	Ключевые слова
1	company, opportunities, people, great, business, management, career, many, culture, employees
2	great, work, company, place, good, people, benefits, environment, cons, lots
3	like, work, working, one, get, best, people, job, years, really
4	good, hours, job, staff, long, pay, work, working, time, get
5	good, work, balance, life, salary, great, environment, place, hours, working

Таблица 4. Темы в негативных отзывах LDA
Table 4. Topics in negative reviews LDA

Тема	Ключевые слова
1	work, people, company, get, good, don't, one, like, management, even
2	work, hours, good, get, working, time, don't, people, job, managers
3	management, work, staff, good, pay, company, people, employees, poor, hours
4	sales, company, gs, oracle, reps, good, bad, ryan, dog, tax
5	good, work, culture, management, great, people, company, leadership, poor, lack

Таблица 5. Темы в позитивных отзывах NMF
Table 5. Topics in positive feedback NMF

Тема	Ключевые слова
1	hours, working, long, people, environment, job, flexible, time, get, opportunities
2	good, work, pay, salary, nice, environment, place, bad, nothing, cons
3	balance, life, work, work-life, growth, learning, culture, less, best, salary
4	company, benefits, best, great, employees, big, opportunities, culture, growth, large
5	great, place, work, people, culture, benefits, none, cons, think, career

Таблица 6. Темы в негативных отзывах NMF
Table 6. Negative messaging themes NMF

Тема	Ключевые слова
1	people, company, employees, great, get, many, don't, like, years, time
2	poor, management, lack, great, pay, culture, senior, support, leadership, colleagues
3	work, life, balance, environment, hours, place, work-life, culture, long, home
4	staff, hours, pay, working, free, managers, food, discount, job, get
5	good, bad, pay, benefits, salary, nothing, culture, management, everything, terrible

Обсуждение результатов тематического анализа

Позитивные отзывы:

• хорошие условия труда: темы, связанные с благоприятной рабочей средой, включают такие слова, как good, great, work, company, people, environment. Это отражает позитивное восприятие сотрудниками условий труда и корпоративной культуры;

• баланс работы и личной жизни: часто встречаются слова balance, life, flexible, hours, что указывает на важность баланса между профессиональными обязанностями и личной жизнью для сотрудников;

• возможности роста: слова opportunities, career, growth, learning подчеркивают наличие возможностей для профессионального развития и карьерного роста, что является значимым фактором удовлетворенности сотрудников.

Негативные отзывы:

• проблемы с руководством: часто упоминаются слова management, poor, lack, leadership, что свидетельствует о наличии проблем в управлении и недостаточной поддержке со стороны руководства;

• низкая оплата труда: слова pay, salary, benefits, nothing указывают на недовольство уровнем заработной платы и социальными льготами, что является одной из основных причин недовольства сотрудников;

• долгие рабочие часы: слова work, hours, long, time, work-life balance отражают перегрузку работой и недостаточный баланс между работой и личной жизнью, что негативно сказывается на удовлетворенности сотрудников.

На основании полученных результатов классификации и тематического анализа можно выделить ключевые факторы, влияющие на удовлетворенность и недовольство сотрудников.

Ключевые факторы удовлетворенности:

– хорошие условия труда: благоприятная рабочая среда, поддержка со стороны коллег и руководства;

– баланс работы и личной жизни: гибкий график, возможность совмещать профессиональные обязанности с личными интересами;

– возможности карьерного роста: наличие программ профессионального развития, возможности для повышения квалификации и продвижения по карьерной лестнице.

Основные причины недовольности:

– проблемы с руководством: недостаточная поддержка, неэффективное управление, отсутствие прозрачности в принятии решений;

– низкая оплата труда: несоответствие уровня заработной платы и социальных льгот ожиданиям сотрудников;

– перегрузка работой: длительные рабочие часы, высокая нагрузка, недостаточный баланс между работой и личной жизнью.

Применение модели DistilBERT и методов тематического моделирования позволило провести детальный анализ отзывов сотрудников, выявить ключевые факторы, влияющие на их удовлетворенность и недовольство. Полученные результаты имеют высокую академическую ценность и могут быть использованы для разработки стратегий управления персоналом, направленных на повышение удовлетворенности и лояльности сотрудников. В дальнейшем рекомендуется расширить исследование, включив дополнительные данные и более сложные модели анализа для повышения точности и глубины понимания факторов, влияющих на удовлетворенность сотрудников.

Практическая значимость и рекомендации

Результаты исследования предоставляют ценные инсайты для оптимизации управления персоналом. Внедрение разработанной модели анализа удовлетворенности сотрудников способствует улучшению различных аспектов HR-менеджмента посредством следующих направлений.

1. Мониторинг настроений сотрудников. Регулярный анализ отзывов сотрудников с использованием автоматизированных методов NLP позволяет оперативно выявлять текущие настроения и потенциальные проблемы внутри организации. Это обеспечивает возможность своевременного реагирования на негативные тенденции, предотвращения ухудшения морального климата и поддержания высокого уровня удовлетворенности персонала.

2. Выявление проблемных областей. Тематический анализ, выполненный с помощью методов LDA и NMF, позволяет идентифицировать конкретные сферы, требующие улучшения. Определение ключевых тем негативных отзывов, таких как проблемы с руководством, низкая оплата труда или перегрузка работой, предоставляет руководству четкое представление о приоритетных областях для вмешательства и улучшений.

3. Разработка целевых стратегий. На основе полученных данных организации могут разрабатывать и внедрять целевые программы, направленные на повышение удовлетворенности сотрудников. Это включает разработку инициатив по улучшению условий труда, баланса между работой и личной жизнью, а также предоставление возможностей для профессионального роста и развития.

4. Улучшение управленческих практик. Результаты анализа подчеркивают важность развития управленческих навыков руководителей. Обучение и развитие менеджеров в области эффективного лидерства, коммуникации и управления конфликтами способствует повышению доверия сотрудников и созданию более позитивной рабочей атмосферы.

5. Оптимизация рабочего времени. Внедрение гибких графиков работы и поддержка удаленного формата труда являются эффективными мерами для улучшения баланса между работой и личной жизнью сотрудников. Такие подходы способствуют снижению стресса, увеличению производительности и общей удовлетворенности сотрудников.

6. Пересмотр системы компенсаций и льгот. Анализ отзывов указывает на значимость конкурентоспособной системы оплаты труда и социальных льгот. Регулярный пересмотр и корректировка уровней заработной платы, а также расширение пакета льгот, соответствующих рыночным стандартам и ожиданиям сотрудников, являются ключевыми факторами мотивации и удержания талантливых специалистов.

Для эффективной реализации предложенных рекомендаций рекомендуется следующая последовательность действий:

1) внедрение системы постоянного мониторинга: автоматизированные инструменты для анализа отзывов должны быть интегрированы в существующие HR-системы для обеспечения непрерывного отслеживания настроений сотрудников;

2) проведение регулярных тематических анализов: периодическое выполнение тематического анализа позволяет актуализировать данные и выявлять новые тенденции и проблемы по мере их возникновения;

3) разработка и реализация программ улучшения: на основе выявленных проблемных областей необходимо разрабатывать конкретные программы и инициативы, направленные на их устранение. Это может включать тренинги для руководителей, пересмотр графиков работы, а также улучшение системы вознаграждений;

4) оценка эффективности внедренных мер: важно регулярно оценивать результаты принятых мер с помощью тех же методов анализа, чтобы определить их влияние на удовлетворенность сотрудников и при необходимости корректировать стратегии.

Применение результатов исследования в управлении персоналом предоставляет организациям мощный инструмент для повышения удовлетворенности и лояльности сотрудников. Автоматизированный мониторинг настроений, тематический анализ проблемных областей и разработка целевых стратегий позволяют принимать обоснованные управленческие решения, способствующие улучшению рабочей среды и достижению стратегических целей компании. В дальнейшем рекомендуется расширить исследование, интегрируя дополнительные источники данных и совершенствуя модели анализа для повышения их точности и адаптивности к динамичным условиям современного рынка труда.

Заключение

1. Разработана и внедрена модель анализа удовлетворенности сотрудников на основе методов глубокого обучения, ориентированная на обработку и интерпретацию текстовых отзывов. Применение модели DistilBERT продемонстрировало высокую точность классификации, особенно в категоризации позитивных и негативных отзывов, что подтверждается показателями Precision и Recall, достигшими 0,81 и 0,87 соответственно для позитивных отзывов. Такой результат свидетельствует о надежности и эффективности модели в контексте многоклассовой классификации данных об удовлетворенности сотрудников.

2. Проведенный тематический анализ с использованием методов LDA и NMF позволил выявить ключевые факторы, влияющие на уровень удовлетворенности и неудовлетворенности персонала. Основные положительные аспекты, такие как благоприятные условия труда, баланс между работой и личной жизнью, а также возможности для карьерного роста, были подтверждены частотным и тематическим анализом. Выявлены критические области, включая проблемы с руководством, низкий уровень оплаты труда и перегрузку работой, что предоставляет ценную информацию для стратегического планирования в управлении человеческими ресурсами.

3. Полученные результаты обладают высокой практической значимостью для организаций, стремящихся к повышению эффективности управления персоналом. Автоматизированный мониторинг настроений сотрудников и выявление проблемных областей на основе анализа текстовых данных позволяют своевременно принимать обоснованные управленческие решения, направленные на улучшение условий труда и повышение лояльности сотрудников. Внедрение разработанной модели будет способствовать не только повышению уровня удовлетворенности сотрудников, но и общей производительности и конкурентоспособности компании.

Список литературы / References

1. Jurafsky D., Martin J. H. (2008) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. USA, Prentice Hall Publ.
2. Devlin J., Chang M.-W., Lee K., Toutanova K. (2019) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
3. Griffiths T. L., Steyvers, M. (2004) Finding Scientific Topics. *Proceedings of the National Academy of Sciences of the United States of America*. 101, 5228–5235.
4. Kaggle. DG. Glassdoor Job Reviews (2021) Available: <https://www.kaggle.com/datasets/davidgauthier/glassdoor-job-reviews> (Accessed 1 October 2024).
5. Blei D. M., Ng A. Y., Jordan M. I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*. 3, 993–1022.
6. Kazinets A. (2024) Application of Deep Learning Methods for Employee Satisfaction Analysis Based on Text Data. *Google Colab*. Available: https://colab.research.google.com/drive/1oFYqwiu2-rty2njxat_y2RX1JCKGhNkKw?usp=sharing (Accessed 1 October 2024).

Поступила 18.11.2024

Принята в печать 16.01.2025

Доступна на сайте 10.07.2025

Received: 18 November 2024

Accepted: 16 January 2025

Available on the website: 10 July 2025

Сведения об авторе

Козинец А. Н., асп. каф. экономики, Белорусский государственный университет информатики и радиоэлектроники

Адрес для корреспонденции

220013, Республика Беларусь,
Минск, ул. П. Бровки, 6
Белорусский государственный университет
информатики и радиоэлектроники
Тел.: +375 17 293-80-46
E-mail: kozinets.science@gmail.com
Козинец Александр Николаевич

Information about the author

Kazinets A. N., Postgraduate at the Department of Economics, Belarusian State University of Informatics and Radioelectronics

Address for correspondence

220013, Republic of Belarus,
Minsk, P. Brovki St., 6
Belarusian State University
of Informatics and Radioelectronics
Tel.: +375 17 293-80-46
E-mail: kozinets.science@gmail.com
Kazinets Aliaksandr Nikolaevich