

МАШИННОЕ ОБУЧЕНИЕ В ОЦЕНКЕ ЦИФРОВОЙ ТРАНСФОРМАЦИИ: АНАЛИЗ СТАТИСТИЧЕСКИХ ДАННЫХ РЕСПУБЛИКИ БЕЛАРУСЬ

Ю. И. Спирина

*студент, институт экономики Нижегородского государственного университета
им. Н. И. Лобачевского, г. Нижний Новгород, Россия, SpirinaIuliia@yandex.ru*

Научный руководитель: А. Л. Сочков

*кандидат технических наук, доцент, Нижегородский государственный
университет им. Н. И. Лобачевского, г. Нижний Новгород, Россия,
an.so2009@yandex.ru*

В данной работе был проведен кластерный анализ показателей по шести областям Республики Беларусь, задействовавший ключевые аспекты цифровой инфраструктуры, инвестиционной активности и доступа к сети-интернет. Данные были предварительно сгруппированы и подвергнуты кластеризации, что позволило выявить различия и проанализировать уровень цифрового неравенства с помощью индекса Джини.

Ключевые слова: типология областей; машинное обучение; анализ больших данных; искусственный интеллект.

MACHINE LEARNING IN THE EVALUATION OF DIGITAL TRANSFORMATION: ANALYSIS OF STATISTICAL DATA OF THE REPUBLIC OF BELARUS

Yu. I. Spirina

*student, institute of economics of Lobachevsky State University of Nizhny Novgorod,
Nizhny Novgorod, Russia, pirinaIuliia@yandex.ru*

Supervisor: A. L. Sochkov

*PhD in technical sciences, associate professor, Lobachevsky State University of Nizhny
Novgorod, Nizhny Novgorod, Russia, an.so2009@yandex.ru*

This paper presents a cluster analysis of indicators for six regions of the Republic of Belarus, focusing on key aspects of digital infrastructure, investment activity, and internet access. The data were pre-grouped and subjected to clustering, which allowed for the identification of differences and the analysis of the level of digital inequality using the Gini index.

Keywords: typology of regions; machine learning; big data analysis; artificial intelligence.

Современные тенденции развития мирового сообщества требуют повсеместной информатизации и автоматизации. Внедрение цифровых технологий меняет не только внутреннюю структуру предприятий и производственных процессов, но и социально-экономическую ситуацию в стране.

Президентом Республики Беларусь объявлена Национальная стратегия цифрового развития, что обуславливает актуальность исследования уровня цифрового неравенства регионов страны. В качестве материала, взяты данные с сайта Национального статистического комитета Республики Беларусь за 2022 год. В выборку вошли шесть показателей областей, которые были предварительно отобраны при помощи матрицы Пирсона: x_1 – общее число организаций ИКТ сектора; x_2 – инвестиции в основной капитал ИКТ сектора; x_3 – использование организациями сети Интернет; x_4 – организации, использовавшие облачные сервисы; x_5 – интернет для взаимодействия с государственными органами; x_6 – доступ интернета для населения.

В качестве метода исследования использован кластерный анализ, хорошо зарекомендовавший себя при изучении объектов типа регионы страны [1–3]. Автоматизированная агломеративная кластеризация, являющаяся одним из иерархических методов распределения, была выполнена при помощи программы. Наилучший результат (оценка произведена методом коэффициента силуэта) был получен при разделении областей Беларуси на 2 кластера (0.779). Для лучшей репрезентативности был взят массив, разделённый на 3 кластера (0.628). В лидеры вышли Гомельская и Минская области; в аутсайдерах остались: Витебская, Гродненская и Могилевская области (таблица).

Средние значения кластеров

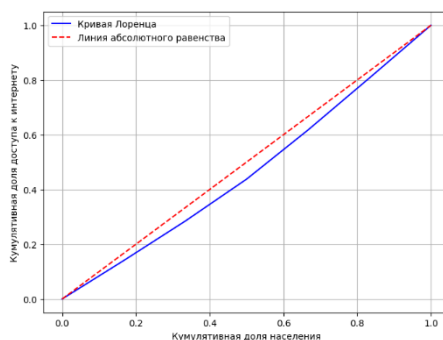
Кластер	x_1	x_2	x_3	x_4	x_5
0	216.0	69.8	98.966667	43.833333	88.066667
1	376.5	81.1	99.750000	35.550000	86.150000
2	263.0	154.7	98.900000	34.400000	86.600000

Методом оценки цифрового неравенства выбран индекс Джини. Подсчёт значений и визуализация были проведены программой, на языке Python по формуле (1):

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2n^2 \bar{x}},$$

где n – количество областей; x_i и x_j – доля населения с доступом к интернету в регионе; $\bar{x}$ – среднее значение доступа.

Полученный индекс $G \approx 0.071$ указывает на низкий уровень цифрового неравенства в доступе к интернет-технологиям в областях Беларуси (рисунок).



Кривая Лоренца

Равномерное распределение информатизации является положительным фактором для социально-экономической стабильности. Однако, существуют значительные различия в уровне цифровой инфраструктуры между областями из выборки. Проведённая кластеризация продемонстрировала значительные расхождения в уровне цифрового развития областей Беларуси. Средние значения кластеров показывают, что регионы с наибольшим числом ИКТ-организаций и высоким уровнем инвестиций, закономерно имеют лучшие показатели по использованию локальных вычислительных сетей и облачных сервисов.

Таким образом, регионы с малыми инвестициями в ИКТ сектор нуждаются в создании специализированных программ по поддержке цифровой экономики и управления внедрением новых технологий. Для успешной цифровизации всех сфер экономической деятельности, требуется особое внимание к развитию цифровых навыков специалистов в тех областях Беларуси, где были зафиксированы низкие показатели использования цифровых технологий.

Библиографические ссылки

1. Перова В. И., Ласточкина Е. И. Нейросетевое моделирование динамики инновационной деятельности в регионах Российской Федерации // Вестн. Ниж. ун-та им. Н. И. Лобачевского. Сер. Соц. науки. 2015. № 3. С. 49–58.
2. Трифонов Ю. В., Сочков А. Л., Миронов Е. А. Типология российских регионов с точки зрения развития человеческого капитала на базе нейросетевого кластерного анализа // Вестн. Ниж. ун-та им. Н. И. Лобачевского. Сер. Соц. науки. 2021. № 4. С. 23–34.
3. Трифонов Ю. В., Сочков А. Л., Соловьев А. Е. Оценка экономического потенциала регионов РФ на основе методологии нейросетевого кластерного анализа // Вестн. Ниж. ун-та им. Н. И. Лобачевского. Сер. Соц. науки. 2021. № 3. С. 38–47.