

УДК 338.001.36

C. Математические и количественные методы
C. Mathematical and Quantitative Methods

АНАЛИЗ ОСОБЕННОСТЕЙ ПОТРЕБИТЕЛЬСКОГО ВЫБОРА В ЦИФРОВОЙ СРЕДЕ

К. И. КОРБУТ¹⁾, О. И. ЛАВРОВА²⁾

¹⁾Леверекс Интернейшнл, ул. Огинского, 6, 220076, г. Минск, Беларусь

²⁾Белорусский государственный университет информатики и радиоэлектроники,
ул. П. Бровки, 6, 220013, г. Минск, Беларусь

Аннотация. Исследуется поведение потребителей интернет-магазина товаров длительного пользования в целях стимулирования продаж продукции премиального сегмента. В качестве ключевой задачи ставится построение прогностической модели, позволяющей с высокой точностью оценить вероятность совершения следующей премиальной покупки на основе аналитики больших данных. В рамках исследования применяются методы многомерного анализа данных, в частности факторный анализ (для выявления латентных паттернов восприятия товара) и кластерный анализ (для группировки пользователей по информационно-поведенческим стратегиям). Устанавливается, что информационная вовлеченность (глубина взаимодействия с отзывами и ценовым фильтром) в наибольшей степени повышает вероятность осуществления премиальной покупки. Результаты модели позволяют реализовать персонализированное отображение контента, сегментировать аудиторию для имейл- и пуш-рассылок, а также учесть региональные особенности и макроэкономические условия, влияющие на предпочтения потребителей. Подтверждается гипотеза о значимости воздействия когнитивных факторов при осуществлении потребительского выбора в цифровой среде.

Ключевые слова: потребительское поведение; логистическая регрессия; премиальная покупка; факторный анализ; кластерный анализ; большие данные; цифровой маркетинг.

Образец цитирования:

Корбут КИ, Лаврова ОИ. Анализ особенностей потребительского выбора в цифровой среде. *Журнал Белорусского государственного университета. Экономика*. 2025;2:4–14. EDN: MQCRIX

For citation:

Korbut KI, Lavrova OI. Analysis of consumer choice features in the digital environment. *Journal of the Belarusian State University. Economics*. 2025;2:4–14. Russian.
EDN: MQCRIX

Авторы:

Кристина Игоревна Корбут – инженер-аналитик.
Ольга Игоревна Лаврова – кандидат экономических наук, доцент; декан инженерно-экономического факультета.

Authors:

Krystsina I. Korbut, analytical engineer.
kristinakorbut1206@gmail.com
Olga I. Lavrova, PhD (economics), docent; dean of the faculty of engineering and economics.
o.lavrova@bsuir.by

ANALYSIS OF CONSUMER CHOICE FEATURES IN THE DIGITAL ENVIRONMENT

K. I. KORBUT^a, O. I. LAVROVA^b

^a*Leverex International, 6 Aginskaga Street, Minsk 220076, Belarus*

^b*Belarusian State University of Informatics and Radioelectronics,
6 P. Browki Street, Minsk 220013, Belarus*

Corresponding author: O. I. Lavrova (o.lavrova@bsuir.by)

Abstract. The article studies the behaviour of consumers of an online store selling durable goods in order to stimulate sales of premium products. The key objective is to build a predictive model that allows for a highly accurate assessment of the likelihood of making the next premium purchase based on big data analytics. The study uses multivariate data analysis methods, in particular, factor analysis (to identify latent patterns of product perception) and cluster analysis (to group users based on information-behavioural strategies). It is established that information involvement (the depth of interaction with reviews and filtering) greatly increases the likelihood of making a premium purchase. The model results make it possible to implement personalised content display, segment the audience for email and push notifications, and take into account regional characteristics and macroeconomic conditions that influence consumer preferences. The hypothesis about the significance of the influence of cognitive factors in making consumer choices in the digital environment is confirmed.

Keywords: consumer behaviour; logistic regression; premium purchase; factor analysis; cluster analysis; big data; digital marketing.

Введение

Современный рынок цифровой торговли характеризуется ростом доли премиального ассортимента, что требует от бизнесменов более осознанного и системного понимания закономерностей и условий выбора, который делает потребитель в интернет-пространстве. Особенno актуальной становится задача выявления групп индивидов, демонстрирующих склонность к совершению дорогих покупок, а также факторов, определяющих такое поведение. В условиях асимметрии информации и избыточности регистрируемых данных, перепроизводства и насыщения рынков при сохранении ограничений потребительских бюджетов руководство компаний стремится к формированию персонализированных стратегий взаимодействия с клиентами, основанных на глубоком анализе цифровых следов.

Важным моментом в прогнозно-аналитической деятельности в рамках исследований потребительских предпочтений является возможность опираться не на субъективные оценки, полученные в результате анализа данных анкет или интервьюирования респондентов, а на информацию, предоставленную в ходе применения инструментов бизнес-аналитики и продвинутых технологий сбора и обработки данных. Одним из наиболее перспективных подходов к анализу и прогнозированию потребительского выбора в цифровой среде является аппарат математической статистики и эконометрики, в частности логистическая регрессия с включением латентных переменных, отражающих информационное и ценовое восприятие товара. Данная модель не только практически реализуема и интерпретируема, но и технически применима в рамках онлайн-витрин, интерфейсов рекомендаций и персонализированных коммуникаций.

Целью исследования выступает выявление значимых факторов, формирующих склонность потребителей к премиальным покупкам, в том числе анализ специфических условий и алгоритмов принятия решений в организованном интернет-пространстве. Предметная область исследования лежит в плоскости цифровой торговли и касается поведенческих паттернов клиентов интернет-магазина, приобретающих товары длительного пользования. Данная проблематика весьма актуальна для развития экономики Республики Беларусь в силу действия фундаментального закона о возвышении общественных потребностей и цифровой трансформации всех сфер общественной жизни.

В основе исследования лежит следующая система гипотез:

- H_1 – гипотеза, предполагающая, что потребители, просматривающие пользовательские отзывы до покупки, с высокой вероятностью выбирают товары премиального сегмента и при последующей покупке сохраняют этот выбор;

- H_2 – гипотеза, подразумевающая, что потребители, характеризующиеся низкой информационной вовлеченностью (по просмотру отзывов, использованию ценового фильтра и другим параметрам), склонны к совершению покупок в массовом сегменте;

- H_3 – гипотеза, согласно которой потребительский выбор в пользу отечественных брендов прямо пропорционально зависит от демографических характеристик (возраст) и территориальной принадлежности (нестоличный регион проживания).

Достижение цели и проверка сформулированных выше гипотез будут осуществляться посредством применения методов многомерного статистического анализа – факторного анализа (для выделения латентных когнитивных и ценовых измерений восприятия товара) и кластерного анализа (для типологизации потребительских стратегий), методов эконометрического моделирования, а также разработки логистической регрессии для оценки вероятности совершения премиальной покупки на индивидуальном уровне. Практическая значимость ожидаемых результатов заключается в получении научно обоснованных рекомендаций по ведению цифрового бизнеса, в частности по проработке решений, связанных с персонализацией торговой витрины и осуществлением эффективных маркетинговых коммуникаций. Проверка гипотезы H_3 дает основания выстраивать стратегии продвижения продукции белорусских производителей с использованием предиктивной аналитики данных цифровых торговых объектов и платформ.

Теоретико-методологические основы анализа потребительского выбора в условиях цифровой среды

Изучение потребительских предпочтений представляет собой междисциплинарное направление, находящееся на стыке экономики, маркетинга, поведенческой психологии и аналитики данных. В условиях цифровизации и широкого распространения технологий сбора и обработки больших данных особую актуальность приобретает анализ поведенческих паттернов, сформированных на основе синтеза эмпирического опыта и базовых теоретических представлений о рациональном потребительском выборе. Большие данные позволяют измерить и учесть когнитивные особенности процесса потребления в целях более осознанного управления им в контексте устойчивого социально-экономического развития.

Фундаментом теоретического анализа служат исследования Д. Канемана и А. Тверски. Их теория перспектив стала поворотным моментом в понимании потребительского поведения в условиях неопределенности: потери ощущаются значимо сильнее, чем эквивалентные выигрыши, а субъективное восприятие вероятностей приводит к систематическим искажениям в потребительской логике [1]. В развитие этой концепции Р. Х. Талер и К. Р. Санстейн предложили механизм подталкивания (*nudge*), демонстрируя, как изменение архитектуры выбора способно корректировать поведение потребителей в желаемом направлении [2].

В рамках современных работ в области цифрового маркетинга развиваются идеи поведенческой и когнитивной экономики на основе системы комплексной аналитики больших данных, собираемых в интернет-пространстве [3]. К. К. Бойер и Дж. Т. М. Халт исследуют поведенческие намерения онлайн-покупателей в зависимости от способа выполнения заказа и качества клиентского опыта [4]. Влияние социальных сетей и алгоритмической архитектуры на структуру решений изучают В. Сундарадж и М. Р. Раджееш [5]. Значение организационного доверия и цифровой инфраструктуры на примере экосистемы *Alibaba* анализируют Дж. Квак, Ю. Чжан и Ц. Ю [6].

В области методов предсказания потребительского поведения заслуживают внимания работы, в рамках которых применяются кликстрик-аналитика и технологии глубокого обучения. Д. Коэн, С. Лессманн и М. Шааль показывают эффективность рекуррентных нейросетей при прогнозировании последовательности действий онлайн-покупателя [7]. В работе Х. Фу и соавторов описывается влияние сенсорного контакта на принятие решений в цифровой среде, что позволяет осмысливать и оценить роль осознания и визуального комфорта [8]. И. О. Паппас и соавторы применяют качественный сравнительный анализ нечетных множеств fsQCA для выявления комбинаций когнитивных и аффективных восприятий, лежащих в основе покупательских намерений [9].

В ряде исследований акцентируется внимание на конкретных условиях, влияющих на потребительский выбор. Дж. Буко, Л. Какалейчик и М. Ференцова выделяют такие факторы потребительского выбора, как удобство, безопасность и субъективное восприятие цены [10]. Ю. Ноцаки и соавторы анализируют поисковые паттерны, сопоставляя их с вероятностью покупки, что важно для оптимизации внутренних алгоритмов интернет-магазина [11]. Системный подход к анализу пользовательских сценариев с помощью последовательностей действий предложили П. Х. Нгуеном и его коллеги в работе [12]. В. Сватосова предлагает перечень значимых факторов, которые должны быть учтены в маркетинговом планировании независимо от масштаба бизнеса для стратегического управления цифровой торговлей [13].

В условиях ограничения доступа к большим данным научные исследования в области потребительского поведения, опирающиеся на методы и инструменты бизнес-аналитики и цифрового маркетинга, ведутся в недостаточном объеме, особенно в отечественном сегменте науки, что требует совершенствования глубины и полноты эмпирической базы, а также разработки соответствующего прогнозно-математического аппарата.

Факторный и кластерный анализ потребительского выбора на основе больших данных

Для проверки поставленных гипотез используется массив транзакционных данных, собранных в белорусском сегменте интернет-пространства в период с 1 января 2024 г. по 1 марта 2025 г. Всего в массив данных включено 12 000 значений, каждое из которых соответствует одной покупке в рамках уникальной

сессии пользователя. Аудиторию торговой онлайн-площадки составляют 58 % женщин и 42 % мужчин. Установлено, что 46 % транзакций совершают пользователи от 25 до 40 лет, 18 % транзакций – пользователи до 25 лет, 30 % транзакций – пользователи от 41 до 60 лет и 6 % транзакций – пользователи старше 60 лет. В разрезе регионов 34 % покупок приходится на г. Минск, 25 % покупок – на Минскую область, а на территории оставшихся пяти областей совершаются от 7 % до 10 % транзакций.

Для каждой транзакции фиксировались следующие параметры:

- *user_id* – уникальный идентификатор пользователя;
- *session_id* – идентификатор сессии;
- *purchase_datetime* – дата и время совершения покупки;
- *category* – категория товара (категории «Смартфоны, ТВ и электроника», «Мебель», «Игрушки» и др.);
- *brand* – производитель товара;
- *product_name* – полное наименование продукта;
- *price_byn* – фактическая цена товара в белорусских рублях;
- *original_price_byn* – цена товара до применения скидки;
- *discount_percent* – процент скидки;
- *rating* – средний пользовательский рейтинг товара;
- *review_count* – общее число отзывов о товаре;
- *review_view_count* – количество просмотров пользователем отзывов перед совершением покупки;
- *review_view_time_sec* – суммарное время (в секундах), затраченное на чтение отзывов;
- *gender* – пол покупателя (F или M);
- *age* – возраст покупателя;
- *region* – регион (область) проживания покупателя (г. Минск, Минская, Гомельская, Витебская, Гродненская или Могилёвская область);
- *filter_price_flag* – бинарный индикатор использования ценового фильтра при навигации.

Помимо перечисленных переменных, для проверки гипотезы о сохранении выбора товаров премиум-класса введены два бинарных признака с учетом пороговых значений стоимости. Признак *premium_flag* (метка премиальности) принимал значение 1, если цена первично приобретенного товара из определенной категории достигала установленного ценового порога или превышала его. Для категории «Смартфоны, ТВ и электроника» ценовой порог составлял 1500 белорусских рублей, для категории «Мебель» – 3000 белорусских рублей, для категории «Бытовая техника» – 2000 белорусских рублей, для категории «Спорт» – 500 белорусских рублей, для категории «Красота и стиль» – 200 белорусских рублей, для категории «Товары для дома» – 1000 белорусских рублей, для категории «Игрушки» – 300 белорусских рублей.

Признак *next_premium_flag* устанавливался равным 1, если в результате следующей по дате транзакции того же пользователя был повторно приобретен премиальный в соответствии с предложенной выше градацией цен товар. Анализ совершенных на торговой онлайн-площадке транзакций показал, что 34 % пользователей приобретали товары премиум-класса, при этом 26 % пользователей при следующей покупке повторили заказ премиального продукта. На рис. 1 представлено распределение премиальных покупок по категориям товаров.

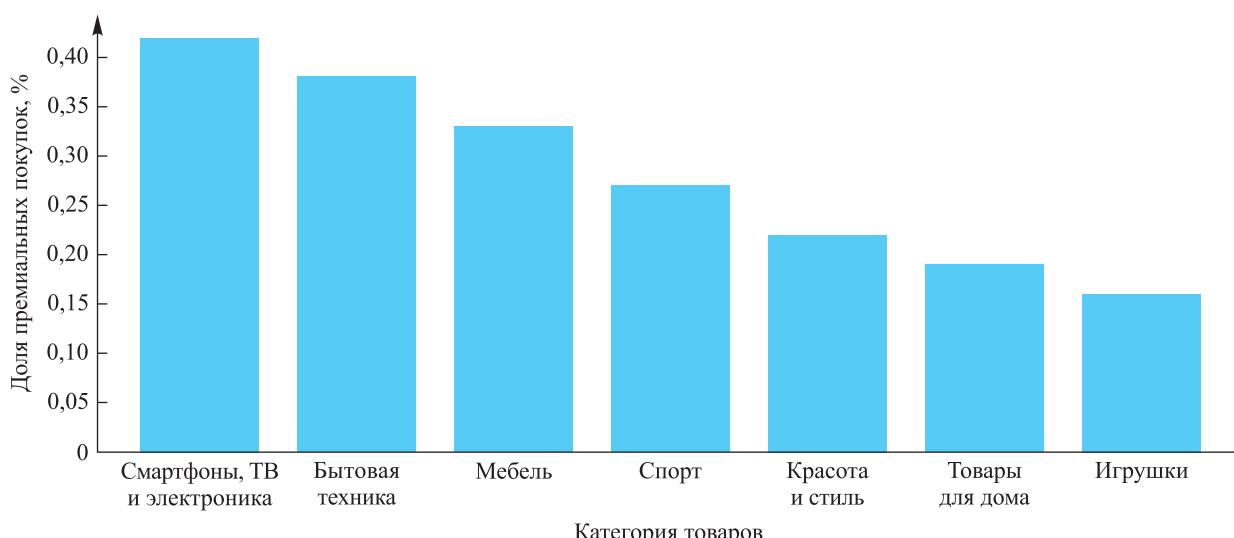


Рис. 1. Распределение премиальных покупок по категориям товаров

Fig. 1. Distribution of premium purchases by product category

Для выявления временных закономерностей на основе параметра *purchase_datetime* были сгенерированы признаки *day_of_week* (день недели) и *hour* (час совершения покупки). Кроме того, с помощью признака *days_from_payday* вычислялась разница между днем покупки и днем заработной платы (условно 15 число месяца). К параметрам *review_view_count* и *review_view_time_sec* был добавлен признак *view_intensity*, равный отношению числа просмотров к общему числу отзывов (*review_count*). Бинарный признак *has_media* указывает на наличие хотя бы одного отзыва или положительного времени просмотра отзывов. Абсолютная величина скидки (признак *discount_abs*) рассчитана как разность между параметрами *original_price_byn* и *price_byn*. Ценовой ранг товара (признак *price_tier*) задавался через квантильное разделение всех цен на пять групп. Для каждого пользователя в агрегированном виде вычислялись общее число покупок (признак *user_total_purchases*), средний чек (признак *user_avg_spend*) и разница в днях между текущей и предыдущей покупкой (признак *days_since_last_purchase*).

В результате был сформирован многомерный массив признаков, подготовленный для проведения факторного и кластерного анализа. В ходе выполнения факторного анализа, направленного на выявление скрытых факторов и уменьшение пространства признаков, исходный набор переменных был разделен на две группы (табл. 1). Интерес представляют группа скрытых факторов «Информационная ценность» (*information value*), которую составляют показатели взаимодействия с отзывами и ценовыми фильтрами, и группа «Качество продукта» (*product quality*), формируемая показателями цены, рейтинга и количества отзывов.

Таблица 1

Группы скрытых факторов и включенные в них переменные

Table 1

Groups of latent factors and variables included in them

Группа скрытых факторов	Наименование переменной	Описание переменной
Информационная ценность	<i>review_view_count</i>	Количество просмотров отзывов
	<i>review_view_time_sec</i>	Суммарное время чтения отзывов
	<i>filter_price_flag</i>	Бинарный индикатор использования ценового фильтра (0 или 1)
	<i>discount_percent</i>	Процент предоставленной скидки
Качество продукта	<i>price_byn</i>	Фактическая цена товара
	<i>rating</i>	Средняя оценка товара
	<i>review_count</i>	Общее количество отзывов о товаре
	<i>premium_flag</i>	Бинарный признак премиального товара (1 – премиальный товар)

Для группы «Информационная ценность» анализ собственных значений показал, что первая компонента факторной модели объясняет примерно 76 % общей дисперсии исходных параметров. Было принято решение ограничиться одной компонентой, получив латентный фактор *factor_info*, который отражает комплексное взаимодействие между вниманием к отзывам и реакцией на ценовые стимулы. Набор факторных нагрузок показал, что наибольший вклад в фактор *factor_info* вносят показатели *review_view_time_sec* и *review_view_count*, а переменная *discount_percent* имеет наименьшую, но все же заметную нагрузку (табл. 2).

Таблица 2

Факторные нагрузки для переменных из группы «Информационная ценность»

Table 2

Factor loadings for variables from the group «Information value»

Наименование переменной	Факторная нагрузка
<i>review_view_time_sec</i>	0,76
<i>review_view_count</i>	0,69
<i>filter_price_flag</i>	0,48
<i>discount_percent</i>	0,36

В группе «Качество продукта» анализ собственных значений выявил две значимые компоненты. Первая компонента *factor_quality_price* объясняет около 55 % общей дисперсии исходных признаков, она преимущественно связана с ценой товара и меткой премиальности, вторая компонента *factor_quality_feedback*, будучи обусловленной в основном рейтингом товара и числом отзывов о нем, объясняет еще около 18 % общей дисперсии. Суммарно две компоненты объясняют порядка 73 % общей дисперсии исходных параметров. Факторные нагрузки показали, что показатели *price_byn* и *premium_flag* имеют высокие коэффициенты корреляции на первой компоненте, тогда как переменные *rating* и *review_count* сильнее загружаются на вторую компоненту.

Для выделения однородных групп потребителей был проведен кластерный анализ на основе поведенческих признаков и трех латентных факторов (*factor_info*, *factor_quality_price*, *factor_quality_feedback*). С помощью оценки коэффициента силуэта было определено оптимальное число кластеров ($k = 4$), позволяющее получить наиболее четкое разделение пользователей в выбранном семипараметрическом пространстве. Все переменные были стандартизованы (приведены к z -оценкам), чтобы исключить влияние разного масштаба измерения и обеспечить корректную работу алгоритма *K-means*. Полученные кластеры представлены в табл. 3.

Таблица 3
Поведенческие кластеры потребителей

Table 3

Consumer behavioural clusters

Название кластера	Описание	Значение фактора <i>factor_info</i>
Ищущие отзывы (<i>review-seekers</i>)	Потребители, демонстрирующие высокую интенсивность просмотра отзывов и длительность времени их чтения. Они тщательно изучают мнения других покупателей перед принятием решения о покупке	Максимальное значение фактора <i>factor_info</i> , которое указывает на глубокий интерес к деталям товара и склонность реагировать на отзывы
Чувствительные к цене (<i>price-sensitive</i>)	Потребители, которые либо практически не просматривают отзывы (низкие значения показателей <i>review_view_count</i> и <i>view_intensity</i>), либо делают это формально, но при этом активно используют ценовой фильтр. Их поведение свидетельствует об ориентации на поиск выгодных предложений и товаров со скидками	Низкое значение фактора <i>factor_info</i> при относительно высоком значении показателя <i>filter_price_flag</i>
Рациональные покупатели (<i>rational comparers</i>)	Потребители, проявляющие высокую активность в использовании ценового фильтра (<i>filter_price_flag</i> = 1) при умеренным интересе к отзывам (средние или невысокие значение показателей <i>view_intensity</i> и <i>review_view_time_sec</i>). Их отличает стремление быстро сопоставить несколько вариантов товара по цене и базовым характеристикам без углубленного изучения отзывов	Значение фактора <i>factor_info</i> находится на среднем уровне (ценовой фильтр потребители из данного кластера используют чаще, чем потребители из кластера «Ищущие отзывы»)
Импульсивные покупатели (<i>impulse buyers</i>)	Потребители, минимально просматривающие отзывы и не использующие ценовой фильтр (<i>filter_price_flag</i> = 0). Они склонны принимать решение о покупке интуитивно или под влиянием внешних факторов (реклама, акции, рекомендации друзей)	Значение фактора <i>factor_info</i> одно из самых низких, что говорит об отсутствии систематичного интереса к деталям товара

В рамках кластера «Ищущие отзывы» доля премиальных покупок оказалась на 18 % выше, чем в среднем по выборке. Более того, если потребитель входит в этот кластер, то вероятность того, что его следующая покупка также окажется премиальной (*next_premium_flag* = 1), возрастает на 12 % по сравнению с остальными кластерами. Пользователи, входящие в кластер «Рациональные покупатели», демонстрируют умеренную вероятность приобретения премиального товара, поскольку они готовы заплатить больше только при условии конкурентоспособной цены и приемлемого соотношения цена – качества у товара. В рамках двух других кластеров доля премиум-транзакций минимальна, что характерно для экономных и экспрессивных потребителей, совершающих дорогостоящие покупки нерегулярно, от случая к случаю, под воздействием эмоций, из-за жестких бюджетных ограничений и по ряду других причин.

Результаты кластерного и факторного анализа согласуются между собой и подтверждают выдвинутые гипотезы H_1 и H_2 . Они использованы для разработки регрессионной модели, предназначеннной для прогнозирования следующей премиальной покупки.

Оценка вероятности покупки премиальных и отечественных товаров на основе логистической регрессии

Разработанная модель логистической регрессии предназначена для прогнозирования вероятности того, что следующая покупка определенного покупателя окажется премиальным товаром. В качестве объясняемой переменной выступает бинарный признак *next_premium_flag*. Данная модель имеет следующий вид:

$$\begin{aligned} \ln \frac{P(\text{next_premium} = 1)}{1 - P(\text{next_premium} = 1)} = & \beta_0 + \beta_1 \text{factor_info} + \beta_2 \text{factor_quality_price} + \\ & + \beta_3 \text{factor_quality_feedback} + \gamma_1 C_1 + \gamma_3 C_3 + \delta_1 \text{price_byn} + \delta_2 \text{discount_percent} + \\ & + \alpha_2 G + \rho_5 \text{region_Minsk} + \mu_2 \text{days_from_payday}, \end{aligned} \quad (1)$$

где C_1 – фиктивная переменная, соответствующая кластеру «Ищущие отзывы»; C_3 – фиктивная переменная, соответствующая кластеру «Импульсивные покупатели»; G – бинарный признак пола (1 – женщина, 0 – мужчина); *region_Minsk* – фиктивная переменная принадлежности к региону (1 – покупатель из г. Минска, 0 – покупатель не из г. Минска).

В качестве объясняющих переменных рассматриваются поведенческий латентный фактор *factor_info*, отражающий степень глубины изучения пользователем отзывов и степень реагирования на скидки; факторы *factor_quality_price* и *factor_quality_feedback*, которые показывают, насколько текущий товар дорогостоящий и премиальный, а также то, какой у него рейтинг и сколько отзывов он имеет. Ценовые и скидочные факторы показывают, что более высокое значение параметра *price_byn* увеличивает шансы покупки товара из премиум-сегмента, а большая скидка снижает их.

Для проверки гипотезы H_3 о выборе отечественных брендов разработана модель

$$\begin{aligned} \ln \frac{P(\text{belarus_brand} = 1)}{1 - P(\text{belarus_brand} = 1)} = & -0,86 + 0,0012 \text{age} - 0,122 \text{discount_percent} + \\ & + 0,052 \text{region_Gomel} + 0,036 \text{region_Grodno} - 0,068 \text{region_Vitebsk}, \end{aligned} \quad (2)$$

где объясняемой переменной является бинарный признак *belarus_brand_flag*, принимающий значение 1, если производитель товара – резидент Республики Беларусь, и принимающий значение 0 во всех прочих случаях. Доля таких транзакций в общем объеме составила 25 %, что свидетельствует о предпочтении отечественных брендов потребителями.

Базовой категорией региона выступает Брестская область, что позволяет интерпретировать остальные коэффициенты относительно нее. В табл. 4 представлены результаты параметризации моделей (1) и (2).

Таблица 4
Оценки параметров моделей и их *p*-значения
Table 4
Estimates of model parameters and their *p*-values

Параметры	Оценка параметров модели (1)	<i>p</i> -Значение для модели (1)	Оценка параметров модели (2)	<i>p</i> -Значение для модели (2)
<i>intercept</i> (β_0)	-1,74	0,001	-0,860 00	0,001
<i>factor_info</i> (β_1)	0,85	0,001	-0,012 00	–
<i>factor_quality_price</i> (β_2)	0,47	0,001	–	–
<i>factor_quality_feedback</i> (β_3)	0,12	0,086	–	–
C_1	0,38	0,001	–	–
C_3	-0,29	0,004	–	–
<i>price_byn</i> (δ_1)	0,22	0,001	0,000 02	–
<i>discount_percent</i> (δ_2)	-0,11	0,028	-0,117 00	0,001
<i>age</i> (α)	0,04	0,045	0,001 10	0,051
$G = 1$ (α_2)	-0,18	0,046	–	–
<i>region_Minsk</i> (ρ_5)	0,31	0,001	-0,011 00	–

Окончание табл. 4
Ending of the table 4

Параметры	Оценка параметров модели (1)	<i>p</i> -Значение для модели (1)	Оценка параметров модели (2)	<i>p</i> -Значение для модели (2)
<i>region_Gomel</i>	–	–	0,048 00	0,057
<i>region_Grodno</i>	–	–	0,033 00	0,071
<i>region_Minsk</i>	–	–	0,074 00	–
<i>region_Vitebsk</i>	–	–	–0,065 00	0,024
<i>days_from_payday</i> (μ_2)	0,02	0,046	–	–

Коэффициент параметра *discount_percent* в обеих логит-моделях отрицателен и статистически значим, что свидетельствует о том, что премиальные и белорусские товары приобретаются в основном без существенных скидок. Демографические факторы обуславливают поведение потребителей следующим образом: мужчины чаще совершают премиальные покупки, а с возрастом вероятность принятия таких решений среди них увеличивается. Временной показатель *days_from_payday* имеет положительную связь с премиальными транзакциями, что отражает необходимость обдумывания в процессе приобретения дорогостоящей продукции. Наибольший вклад в совершение премиальной покупки вносит поведенческий фактор информационной вовлеченности *factor_info*: его увеличение на одну стандартную единицу повышает шансы покупки в 2,3 раза. Следует отметить, что кластерный анализ потребителей позволил усилить объясняющую способность модели: пользователи из кластера «Ищущие отзывы» демонстрируют вероятность премиальной покупки выше среднего уровня, тогда как для пользователей из кластера «Импульсивные покупатели» характерна обратная тенденция. Это обстоятельство подтверждает устойчивые поведенческие различия между сегментами.

Согласно результатам оценивания логит-модели (2) выбор отечественного товара зависит от возраста потребителя и региона его проживания. Вероятность выбора белорусского бренда выше среди покупателей старше 40 лет, а также среди жителей Гомельской и Гродненской областей. Информационный фактор *factor_info* не оказывает влияния на выбор отечественной продукции, поскольку она либо традиционно узнаваема, либо рекомендации по ее приобретению потребитель получает из других коммуникационных источников. Данный результат свидетельствует о наличии паттернов локального выбора, обусловленных демографическими и территориальными признаками.

Анализ эффективности логит-моделей. Прогнозирование

Для анализа эффективности разработанных моделей логистической регрессии были рассчитаны показатели, характеризующие их общее качество и прогнозные способности (табл. 5). Так, для обучающей и тестовой выборок были построены ROC-кривые, на основе анализа которых было установлено, что модели (1) и (2) в 78 % и 72 % случаев соответственно корректно различают положительные (премиальные, отечественные) и отрицательные (непремиальные, импортные) исходы, что особенно важно в условиях несбалансированной выборки.

Таблица 5

Оценка качества моделей логистической регрессии

Table 5

Evaluation of the quality of logistic regression models

Показатели	Модель (1)	Модель (2)
AUC ROC	0,78	0,75
<i>precision</i> (5 %)	0,35	0,32
<i>recall</i>	0,51	0,41
<i>F</i> -оценка	0,50	0,45
Прогнозируемая вероятность, %	55–60	30–32
Наиболее значимые факторы	<i>factor_info</i> , <i>factor_quality_price</i>	<i>discount_percent</i> , <i>region_Gomel</i> , <i>region_Grodno</i> , <i>region_Vitebsk</i>

Показатель *precision* (5 %), равный 0,35, означает, что среди верхних 5 % пользователей по прогнозной вероятности около 35 % пользователей действительно совершают премиум-покупку. Показатель полноты *recall* при пороговом значении 0,5 достиг 0,51, что указывает на способность модели обнаруживать около 51 % реальных покупателей премиум-товаров. Значение *F*-оценки 0,50 подтверждает сбалансированное соотношение между показателем полноты *recall* и показателем точности *precision* для модели (1).

Поскольку модель (1) оказалась более высокого качества, она была использована для расчетов индивидуальных прогнозных вероятностей следующей премиальной покупки на основе формулы

$$P(\text{next_premium} = 1) = \frac{1}{1 + e^{-z}}, \quad (3)$$

где в качестве z принимались значения, полученные в результате оцененной логистической регрессии по тестовой выборке данных ($n = 2400$). В табл. 6 представлено распределение вероятностей премиальной и обычной покупки по поведенческим кластерам.

Таблица 6
Распределение вероятностей премиальной и обычной покупки
по кластерам

Table 6

Probability distribution of premium and regular purchases by cluster

Название кластера	Средняя $P(\text{next_premium_flag} = 1)$	Медиана	Медиана $1 - P$	Стандартное отклонение
Ищущие отзывы	0,47	0,48	0,52	0,14
Рациональные покупатели	0,36	0,34	0,66	0,15
Чувствительные к цене	0,28	0,27	0,73	0,13
Импульсивные покупатели	0,21	0,20	0,80	0,12

На рис. 2 представлено распределение предсказанных вероятностей следующей премиальной покупки по поведенческим кластерам. По ней можно оценить медианные значения, межквартильный размах и выбросы в распределении вероятностей внутри каждой группы.

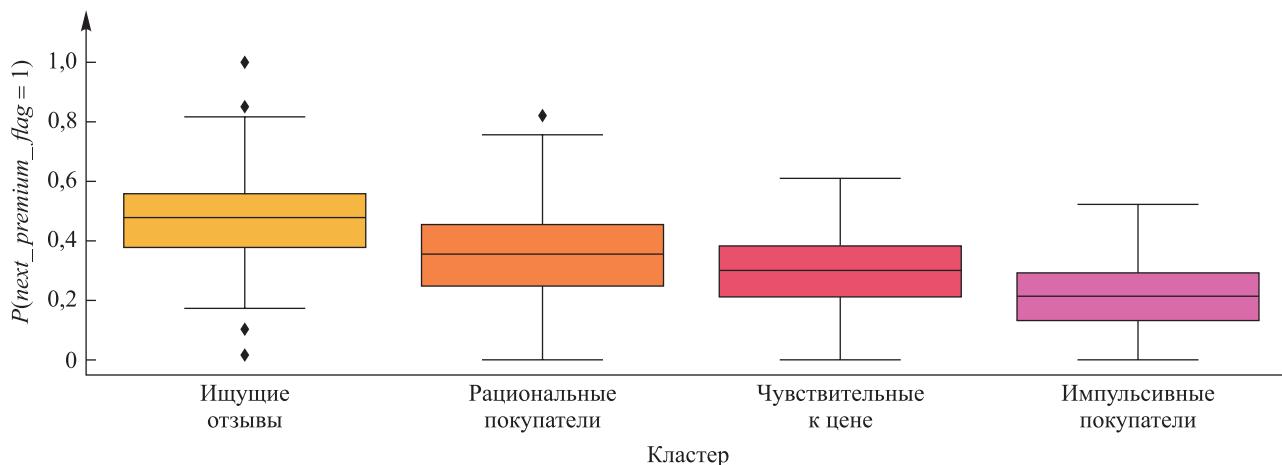


Рис. 2. Распределение предсказанных вероятностей следующей премиальной покупки
по поведенческим кластерам

Fig. 2. Distribution of predicted probabilities of the next premium purchase by behavioural clusters

У покупателей из кластера «Ищущие отзывы» медианная вероятность выше, что подтверждает гипотезу о более высокой склонности таких пользователей к премиальным покупкам. Кластер «Импульсивные покупатели», напротив, характеризуется более низкими вероятностями (часто меньше 0,2). Два других кластера демонстрируют промежуточные значения с широким разбросом.

Результаты моделирования позволяют подтвердить гипотезу H_2 о связи между низкой информационной вовлеченностью и склонностью к массовому потреблению. В модели логистической регрессии

фактор информационной вовлеченности *factor_info* имеет положительный и статистически значимый коэффициент ($\beta = 0,85$; $p < 0,001$), что означает экспоненциальное уменьшение вероятности премиальной покупки при снижении значений данного фактора. Соответственно, чем ниже значение фактора *factor_info*, тем выше вероятность выбора товаров из массового сегмента.

Особенно отчетливо это наблюдается на примере поведенческого кластера «Импульсивные покупатели», в рамках которого информационная вовлеченность минимальна, а медианное значение $P(next_premium_flag = 1)$ составляет всего 0,20. Следовательно, вероятность обычной покупки $1 - P$ среди потребителей, входящих в данный кластер, составляет 0,80, что фактически отражает доминирование недорогих покупок из массового сегмента. Аналогичные значения характерны для кластера «Чувствительные к цене», где вероятность $1 - P$ составляет 0,73.

Заключение

Результаты многомерного статистического и регрессионного анализа потребительского выбора, проведенного на основе больших данных, которые были собраны на белорусской площадке цифровой торговли, позволили проверить ряд гипотез о предпочтениях в покупках потребителей в условиях интернет-пространства. Исследовано влияние поведенческих, ценовых и демографических факторов на принятие решений о покупке как товаров из премиального и массового сегментов, так и товаров из специальных групп, в частности продукции отечественных брендов. Разработанные логит-модели позволяют с высокой точностью прогнозировать вероятность определенной покупки, что открывает широкие возможности для их применения в системах персонализированного маркетинга, интерфейсах цифровых витрин и коммуникационных каналах субъектов цифрового бизнеса. Так, потребителям с высоким уровнем информационной вовлеченности целесообразно демонстрировать предложения из премиального сегмента с применением инструментов, усиливающих когнитивную активность, например инструментов расширенного визуального сопровождения, сравнительных таблиц, галерей и др. Разделение на поведенческие кластеры позволяет осознанно реализовывать сегментированную коммуникацию: предлагать покупателям с высокой вероятностью премиальной покупки обзоры новинок, брендовые предложения и отображать для остальных покупателей краткие описания, товары со скидками и информацию о специальных условиях покупки.

На основе модели вероятности выбора белорусского товара выделяются другие факторы и практические подходы. Установлено, что наибольшее влияние на выбор отечественного бренда оказывают регион проживания и возраст покупателя. В связи с этим необходимо проводить региональную и возрастную адаптацию маркетинговых стратегий. Рекомендуется визуально выделять товары белорусского производства в интерфейсе цифровых торговых объектов, использовать патриотическую символику с указанием происхождения и применять культурные маркеры («Сделано рядом», «Надежно и понятно»). Целесообразно исключать избыточные когнитивные стимулы, поскольку выявлено, что информация о скидках, отзывах и сравнении с другими брендами не оказывает существенного влияния на вероятность выбора товара белорусского бренда. Выявленная закономерность указывает на преимущественную эмоционально-культурную мотивацию потребительского выбора и требует соответствующего тона в коммуникации.

Однако важно учитывать общую макроэкономическую ситуацию: даже при высокой индивидуальной склонности к совершению премиальной покупки потребитель может приобретать недорогие товары под воздействием внешних негативных условий (финансовой нестабильности, кризисных проявлений в экономике, эффекта толпы). Полученный результат подчеркивает необходимость регулярной адаптации и совершенствования моделей с учетом новых данных о предпочтениях потребителя в более широком социально-экономическом контексте.

Библиографические ссылки

1. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica*. 1979;47(2):263–292. DOI: 10.2307/1914185.
2. Thaler RH, Sunstein CR. *Nudge. Improving decisions about health, wealth, and happiness*. New Haven: Yale University Press; 2008. 293 p.
3. Фролова ДА, Гулецкая ЕА. Использование больших данных для повышения лояльности к бренду. В: Богуш ВА, редактор. *Big data и анализ высокого уровня. Сборник научных статей X Международной научно-практической конференции; 13 марта 2024 г.; Минск, Беларусь. Часть 1*. Минск: Белорусский государственный университет информатики и радиоэлектроники; 2024. с. 181–186. EDN: JZMYHН.
4. Boyer KK, Hult GTM. Customer behavioral intentions for online purchases: an examination of fulfillment method and customer experience level. *Journal of Operations Management*. 2006;24(2):124–147. DOI: 10.1016/j.jom.2005.04.002.

5. Sundararaj V, Rejeesh MR. A detailed behavioral analysis on consumer and customer changing behavior with respect to social networking sites. *Journal of Retailing and Consumer Services*. 2021;58:102190. DOI: 10.1016/j.jretconser.2020.102190.
6. Kwak J, Zhang Yu, Yu J. Legitimacy building and e-commerce platform development in China: the experience of Alibaba. *Technological Forecasting and Social Change*. 2019;139:115–124. DOI: 10.1016/j.techfore.2018.06.038.
7. Koehn D, Lessmann S, Schaal M. Predicting online shopping behaviour from clickstream data using deep learning. *Expert Systems with Applications*. 2020;150:113342. DOI: 10.1016/j.eswa.2020.113342.
8. Fu H, Manogaran G, Wu K, Cao M, Jiang S, Yang A. Intelligent decision-making of online shopping behavior based on Internet of things. *International Journal of Information Management*. 2020;50:515–525. DOI: 10.1016/j.ijinfomgt.2019.03.010.
9. Pappas IO, Kourouthanassis PE, Giannakos MN, Chrissikopoulos V. Explaining online shopping behavior with fsQCA: the role of cognitive and affective perceptions. *Journal of Business Research*. 2016;69(2):794–803. DOI: 10.1016/j.jbusres.2015.07.010.
10. Bucko Jo, Kakalejčík L, Ferencová M. Online shopping: factors that affect consumer purchasing behaviour. *Cogent Business & Management*. 2018;5(1):1535751. DOI: 10.1080/23311975.2018.1535751.
11. Nozaki Yu, Watanabe F, Satoh T. Analysis of item selection behavior in online shopping. In: Indrawan-Santiago M, Pardede E, Salvadori IL, Steimbauer M, Khalil I, Anderst-Kotsis G, editors. *iiWAS2018. Proceedings of the 20th International conference on information integration and web-based applications & services; 2018 November 19–21; Yogyakarta, Indonesia*. New York: Association for Computing Machinery; 2018. p. 41–45.
12. Nguyen PH, Turkay C, Andrienko G, Andrienko N, Thonnard O, Zouaoui J. Understanding user behaviour through action sequences: from the usual to the unusual. *IEEE Transactions on Visualization and Computer Graphics*. 2018;25(9):2838–2852. DOI: 10.1109/tvcg.2018.2859969.
13. Svatosova V. The importance of online shopping behavior in the strategic management of e-commerce competitiveness. *Journal of Competitiveness*. 2020;12(4):143–160. DOI: 10.7441/joc.2020.04.09.

Статья поступила в редакцию 19.08.2025.

Received by editorial board 19.08.2025.