

АЛГОРИТМИЗАЦИЯ АНАЛИЗА СПРОСА В УСЛОВИЯХ МНОГОМЕРНОСТИ ДАННЫХ: РАЗВИТИЕ МЕТОДИКИ А. А. КОЛЯДЫ

Галин Евгений Борисович магистр экономических наук, слушатель Евразийской школы менеджмента и администрирования, г. Минск

УДК 005.21

Аннотация

Целью исследования является разработка теоретико-методологического подхода к алгоритмизации анализа рыночного спроса в условиях высокой размерности и гетерогенности данных на этапе Market Value and Complexity-1 (MVC-1) методики бизнес-моделирования А. А. Коляды. Методологической основой выступает теория мультиуровневой природы бизнес-модели, дополненная инструментарием сокращения признаков пространства. В работе предлагается формализованный алгоритмический конвейер, комбинирующий экспертную, энтропийную и модель-ориентированную оценку значимости признаков, а также вводится параметр — управленческий порог детализации ε — как механизм контроля глубины анализа. Работоспособность подхода иллюстрируется на условном примере рынка долевого жилищного строительства. Показано, что предлагаемый подход концептуально обеспечивает воспроизводимое сокращение размерности признаков пространства без существенной потери аналитической информативности и расширяет возможности практического применения методики А. А. Коляды в корпоративной и регуляторной аналитике. Основной вклад работы носит методологический характер и состоит в интеграции инструментов Data Science в структуру мультиуровневого бизнес-моделирования.

Ключевые слова: алгоритмизация спроса; методика Коляды; мультиуровневость; сокращение размерности; порог детализации; ключевые характеристики спроса.

Abstract

The aim of the study is to develop a theoretical and methodological approach to the algorithmization of market demand analysis under conditions of high dimensionality and data heterogeneity at the Market Value and Complexity-1 (MVC-1) stage of A. A. Kolyada's business-modeling methodology. The methodological basis is the theory of the multi-level nature of the business model, supplemented by feature-space reduction tools. The paper proposes a formalized algorithmic pipeline combining expert, entropy-based, and model-oriented feature significance assessment, and introduces a parameter — the managerial detail threshold ε — as a mechanism for controlling analytical depth. The feasibility of the approach is illustrated using a conditional example of the shared-equity housing construction market. It is shown that the proposed approach conceptually ensures a reproducible reduction of the feature space dimensionality without significant loss of analytical value and expands the practical applicability of Kolyada's methodology in corporate and regulatory analytics. The main contribution is methodological and consists in integrating Data Science tools into the structure of multi-level business modeling.

Keywords: demand algorithmization; Kolyada methodology; multi-levelness; dimensionality reduction; detail threshold; key demand characteristics.

Введение

Современная цифровая экономика характеризуется быстрым ростом объема, скорости генерации и размерности данных, порождаемых поведением потребителей и рыночной инфраструктурой. В этих условиях традиционные подходы к анализу спроса — опросные методики, экспертные оценки и классические линейные эконометрические инструменты — демонстрируют ограниченную адекватность. Они не всегда способны корректно фиксировать нелинейные зависимости, своевременно обнаруживать слабые сигналы рынка и эффективно обрабатывать большие и неоднородные массивы данных без потери информативности.

Эта проблема особенно значима для капиталоемкого и конкурентного сектора первичного жилья, где спрос чувствителен к макроэкономическим изменениям и формируется под влиянием сложного сочетания экономических, социальных и поведенческих факторов. Даже небольшие сдвиги в ключевых параметрах, таких как ипотечные ставки, способны вызывать непропорциональные изменения в поведении покупателей.

В этой связи развитие инструментов, обеспечивающих алгоритмизированное, воспроизводимое и управляемое выделение ключевых характеристик спроса, представляет собой актуальную теоретико-методологическую задачу. Универсальная методика бизнес-моделирования, предложенная А. А. Колядой, обладает потенциалом для такого развития благодаря ориентации на мультиуровневую природу бизнес-модели и процедуру формирования рыночных кластеров спроса на этапе Market Value and Complexity-1 (далее MVC-1) [1; 2]. Однако практическая реализация методики в цифровой среде сталкивается с «проклятием размерности»: количество потенциально релевантных признаков резко возрастает, что делает ручной экспертный отбор трудоёмким и недостаточно устойчивым.

Целью исследования является разработка теоретико-методологического подхода — формализованного алгоритмического конвейера для оценки важности признаков и фильтрации признакового пространства, обеспечивающего воспроизводимый и интерпретируемый набор ключевых характеристик спроса (Key Demand Characteristics, далее KDC) для последующей кластеризации и стратегического анализа.

Для достижения цели поставлены следующие задачи:

1. концептуализировать формализованный алгоритмический конвейер, комбинирующий экспертные и объективные методы оценки важности признаков [3; 4];
2. ввести и формализовать управленческий порог детализации ε как механизм баланса между глубиной аналитики и практической применимостью модели;
3. описать способ интеграции конвейера в этап MVC-1 методики А. А. Коляды;
4. проиллюстрировать логику подхода на условном примере рынка долевого строительства и сформулировать направления его эмпирической верификации.

Исследовательский вопрос формулируется так: способен ли формальный алгоритм оценки значимости признаков с последующей фильтрацией по порогу ε обеспечить объективное и воспроизводимое выделение ограниченного набора KDC, достаточного для построения интерпретируемых рыночных кластеров при снижении вычислительной и когнитивной нагрузки аналитического процесса?

Научная новизна заключается в концептуальной интеграции методов машинного обучения (оценка важности признаков, кластеризация) в структуру методики Коляды, а также в формальном введении параметра ε как управляемого механизма сокращения размерности признакового пространства, что расширяет возможности мультиуровневого подхода в условиях больших данных.

Теоретическая рамка и обзор литературы

Концепция бизнес-модели остаётся одной из центральных в стратегическом менеджменте. Классические работы описывают бизнес-модель как способ создания, доставки и удержания ценности [5; 6]. Обзорные исследования систематизируют направления инновации бизнес-модели и подчёркивают роль данных как источника её обновления [7; 8; 9]. Отдельное внимание уделяется влиянию цифровых и динамических способностей на трансформацию бизнес-моделей [10].

Методика А. А. Коляды развивает эти представления через теорию мультиуровневой природы бизнес-модели, в которой этап MVC-1 отвечает за формирование кластеров целевого рынка [1; 2; 11]. Слабым местом ручной реализации этого этапа является зависимость от экспертного отбора признаков, что снижает воспроизводимость результатов при росте размерности данных.

В области анализа данных задача сокращения размерности решается методами отбора признаков (feature selection), среди которых выделяют фильтрующие (энтропия, взаимная информация по К. Шеннону [12]) и обёрточные/встроенные подходы,

использующие важность признаков в моделях машинного обучения. Метод SHAP (SHapley Additive exPlanations) [13] обеспечивает теоретически обоснованное аддитивное распределение вклада признаков. Для сегментации применяется алгоритм k-means [14]. Однако эти инструменты, применяемые изолированно, нередко утрачивают связь с управленческой интерпретируемостью. Настоящая работа предлагает восполнить этот разрыв, встроив инструменты Data Science в управленческую методологию с сохранением интерпретируемости через параметр ε .

Предлагаемый методологический подход

Работа носит теоретико-методологический характер. Предлагаемый алгоритмический конвейер включает восемь логически связанных шагов. Приводимые далее числовые значения носят иллюстративный характер и служат демонстрации логики работы подхода; их эмпирическая верификация на репрезентативной выборке является предметом отдельного исследования.

Шаг 1. Формализация исходного признакового пространства (F). Для рынка долевого строительства определяется исходное многомерное пространство из пяти групп признаков с дискретными уровнями: цена за м² (3 уровня), общая площадь (4 уровня), отделка (2 уровня), количество комнат (5 уровней), этажность (3 уровня). Полное пространство как декартово произведение уровней составляет $3 \times 4 \times 2 \times 5 \times 3 = 360$ гипотетических профилей жилья.

Шаг 2. Нормализация и кодирование. Для приведения разнородных показателей к сопоставимому виду применяются: Min-Max нормализация $x' = (x - \min(x)) / (\max(x) - \min(x))$; логарифмическое преобразование с последующей Min-Max нормализацией для правосторонне скошенных распределений (цены, ипотечные выплаты); инверсное преобразование $x' = (\max(x) - x) / (\max(x) - \min(x))$ для признаков по принципу «меньше — лучше». Категориальные признаки кодируются порядковым или бинарным способом.

Шаг 3. Оценка важности признаков. Вклад каждого признака f_i оценивается комбинацией трёх методов для минимизации субъективности:

а) *Экспертный метод* задаёт стартовые веса w_i^e на основе методологии Коляды и опроса аналитиков.

б) *Энтропийный метод* оценивает информативность признака через нормированную энтропию Шеннона. Для признака с распределением по уровням $\{p_j\}$ энтропия $H(f_i) = - \sum_j p_j \log_2 p_j$; информативная значимость определяется как $w_i^{ent} = 1 - H(f_i) / \log_2(k_i)$, где k_i — число уровней признака.

в) *Модель-ориентированный метод* использует усреднённые абсолютные значения SHAP по обученной модели градиентного бустинга, прогнозирующей индикатор потребительского выбора: $w_i^{shap} = (1/n) \sum_n |\phi_{i,n}|$, где $\phi_{i,n}$ — вклад признака f_i для наблюдения n .

Итоговый вес признака рассчитывается как средневзвешенное трёх оценок: $w_i = \alpha \cdot w_i^e + \beta \cdot w_i^{ent} + \gamma \cdot w_i^{shap}$, где $\alpha + \beta + \gamma = 1$ (в базовой конфигурации $\alpha = \beta = \gamma = 1/3$). Нормализованная важность вычисляется относительно наиболее значимого признака: $N(f_i) = w_i / \max(w)$.

Шаг 4. Фильтрация ключевых характеристик спроса (KDC) по порогу ε . Вводится управленческий порог детализации $\varepsilon \in [0, 1]$. В итоговую модель включаются только признаки, чья нормализованная важность превышает порог:

$$f_i \in KDC \Leftrightarrow N(f_i) > \varepsilon.$$

Параметр ε — механизм управления сложностью модели: высокое значение оставляет «ядро» спроса, низкое — сохраняет уточняющие, но менее значимые факторы. В отличие от статистического порога отсекающего, ε задаётся лицом, принимающим решения, исходя из требуемого баланса детализации и управляемости.

Шаг 5. Построение сокращённой модели и кластеризация (MVC-1). На основе набора KDC формируется оптимизированное признаковое пространство, к которому применяется алгоритм k-means [14] для выделения однородных групп потребительских

запросов. Это формализует этап MVC-1 методики Коляды. Оптимальное число кластеров k определяется методом «локтя» и максимумом среднего коэффициента силуэта.

Шаг 6. Расчёт интегрального показателя привлекательности ЖК (A_i). Для ранжирования объектов вводится показатель, агрегирующий два экономических фактора — нормированную цену за м² ($\tilde{P}$) и нормированную ставку ипотеки ($\tilde{R}$):

$$A_i = \omega_1 \cdot \tilde{P}_i + \omega_2 \cdot \tilde{R}_i,$$

где $\omega_1 = 0,6$, $\omega_2 = 0,4$ — веса приоритетов покупателя; значения нормированы по принципу «меньше — лучше». Показатель A_i рассматривается не как независимая верификация оценки важности, а как производный управленческий инструмент ранжирования, операционализирующий выявленное доминирование экономических факторов.

Шаг 7. Процедура валидации. Для оценки устойчивости результатов предлагается двухкомпонентная процедура: (1) бутстрэп-анализ с генерацией случайных подвыборок и повторным выполнением шагов 3–5, с оценкой стабильности состава KDC по индексу Жаккара и воспроизводимости кластерных профилей; (2) анализ чувствительности к весам ω_1 , ω_2 путём их варьирования в пределах $\pm 20\%$ для оценки стабильности рейтинга.

Шаг 8. Программная реализация. Конвейер реализуем в среде Python 3.9 с использованием библиотек pandas, scikit-learn, scipy, catboost, shap.

Модельный пример

Для демонстрации логики подхода приведём условный расчёт на данных рынка долевого строительства г. Минска и Минского района (иллюстративные значения).

Применение конвейера к признаковому пространству даёт ранжированную иерархию значимости (таблица 1).

Таблица 1. Нормализованная важность групп признаков (иллюстративный пример)

Группа признаков	N(f _i)	Статус в модели
Цена за м ²	1,000	Ядро спроса
Общая площадь (метраж)	0,313	Ключевой уточняющий фактор
Отделка	0,250	Условно-значимый фактор
Количество комнат	0,040	Второстепенный фактор
Этажность	0,017	Второстепенный фактор

Пример иллюстрирует доминирование ценового фактора. Метраж и тип отделки вносят заметный, но существенно более слабый вклад. Низкая значимость «количества комнат» концептуально объясняется его высокой коллинеарностью с метражом; при этом следует учитывать, что SHAP-оценки при выраженной мультиколлинеарности могут нестабильно распределять вклад между коррелирующими признаками — данное ограничение адресуется процедурой валидации (шаг 7).

Сценарный анализ выбора ε представлен в таблице 2.

Таблица 2. Влияние порога ε на состав и сложность модели (иллюстративный пример)

Порог ε	Включённые признаки	Число комбинаций	Интерпретация
0,50	Цена	3	Избыточное упрощение
0,30	Цена + Метраж	12	Баланс, но без качественного параметра
0,25	Цена + Метраж + Отделка	24	Оптимальный компромисс
0,05	+ Комнаты	120	Избыточная детализация
—	Все признаки	360	Исходная размерность

Значение $\varepsilon = 0,25$ иллюстрирует компромисс: сохраняются три доминирующих фактора, а пространство сокращается с 360 до 24 комбинаций.

Кластеризация k-means на реальных потребительских запросах в сокращённом пространстве KDC выделяет три интерпретируемых сегмента (таблица 3).

Таблица 3. Профили кластеров спроса (иллюстративный пример)

Кластер	Цена (\$/м ²)	Площадь (м ²)	Отделка
1. Рациональный спрос	≈ 1300	до 45	черновая
2. Семейный комфорт	1300–1500	46–70	базовая

3. Статусное жильё	> 1500	> 70	чистовая
--------------------	--------	------	----------

Показатель A_i операционализирует ранжирование объектов: в верхней части рейтинга оказываются проекты с благоприятным сочетанием невысокой цены и конкурентных ипотечных условий, в нижней — премиальные объекты со стандартными условиями финансирования.

Обсуждение

Предлагаемый подход концептуально решает исходную задачу: обеспечивает воспроизводимое и управляемое сокращение признакового пространства с интеграцией результатов в этап MVC-1 методики Коляды. Главная ценность состоит в системной интеграции инструментов Data Science в управленческую методологию и формализации параметра ε как механизма контроля уровня детализации.

Интерпретация модельного примера согласуется с положениями поведенческой экономики: доминирование цены отражает действие бюджетного ограничения, умеренная значимость метража сопряжена с ростом популярности функциональных планировок, а низкая значимость числа комнат обусловлена корреляцией с площадью. Преимущество предлагаемого механизма перед чисто статистическими процедурами отбора признаков состоит в сохранении управленческой интерпретируемости через порог ε и в способности интегрировать данные из разнородных источников.

Теоретическая значимость заключается в развитии концепции мультиуровневой бизнес-модели и демонстрации пути её операционализации для задач продуктового рыночного анализа с использованием алгоритмов машинного обучения.

Практические импликации. Для девелоперов – смещение акцента продуктовой политики в сторону оптимизации ценообразования и партнёрских ипотечных предложений, интеграция конвейера в CRM-системы для группировки лидов. Для банков – развитие кластер-ориентированных ипотечных продуктов и использование A_i как дополнительного признака скоринга. Для регуляторов – построение платформы мониторинга динамики показателей по кластерам.

Ограничения и направления развития. Работа носит концептуальный характер, а приведённые числовые значения иллюстративны; их полноценная эмпирическая верификация требует репрезентативной выборки и оценки качества прогностической модели. Дальнейшие направления: адаптация методики ко вторичному рынку жилья; разработка динамической версии алгоритма для отслеживания временных трендов; интеграция методов обработки естественного языка (NLP) для учёта неструктурированных источников (отзывы, новости) и выявления слабых сигналов рынка.

Заключение

Предложен теоретико-методологический подход к алгоритмизации этапа MVC-1 методики А. А. Коляды, основанный на формализованной оценке важности признаков и фильтрации по управленческому порогу детализации ε . Подход концептуально решает проблему высокой размерности признакового пространства и формализует переход от множества потенциальных характеристик к ограниченному набору ключевых факторов (KDC), релевантных для стратегических решений. Основной вклад работы носит методологический характер и состоит в интеграции инструментов Data Science в структуру мультиуровневого бизнес-моделирования с сохранением управленческой интерпретируемости. Предложенный подход обладает прикладной перспективой для девелоперов, банков и регуляторов и задаёт основу для дальнейшей эмпирической, динамической и многоисточниковой верификации.

Список литературы

1. Коляда А. А. Теория мультиуровневой природы (мультиуровневости) бизнес-модели, бизнес-моделирование и инновация бизнес-модели в стратегическом менеджменте // Вестник евразийской науки. 2024. Т. 16, № 4. С. 1–18. DOI: 10.15862/60ECVN424.
2. Коляда А. А., Плехова Ю. О. Определение понятия бизнес-модели и требований к нему в целях стратегического менеджмента организации // Вестник Нижегородского

университета им. Н. И. Лобачевского. Серия: Социальные науки. 2024. № 1(73). С. 7–16. DOI: 10.52452/18115942_2024_1_7.

3. Sorescu A. Data-driven business model innovation // *Journal of Product Innovation Management*. 2017. Vol. 34, no. 5. P. 691–696. DOI: 10.1111/jpim.12398.

4. Коляда А. А., Плехова Ю. О. Методика разработки бизнес-модели организации и прогнозирования её экономической эффективности // *Теория и практика общественного развития*. 2023. № 8(184). С. 101–112. DOI: 10.24158/TIPOR.2023.8.12.

5. Osterwalder A., Pigneur Y. *Business Model Generation*. — Hoboken: John Wiley & Sons, 2010. — 288 p. — ISBN 978-0-470-87641-1.

6. Teece D. J. Business models, business strategy and innovation // *Long Range Planning*. 2010. Vol. 43, no. 2–3. P. 172–194. DOI: 10.1016/j.lrp.2009.07.003.

7. Foss N. J., Saebi T. Fifteen years of research on business model innovation // *Journal of Management*. 2017. Vol. 43, no. 1. P. 200–227. DOI: 10.1177/0149206316675927.

8. Geissdoerfer M., Vladimirova D., Evans S. Sustainable business model innovation: A review // *Journal of Cleaner Production*. 2018. Vol. 198. P. 401–416. DOI: 10.1016/j.jclepro.2018.06.240.

9. Ramdani B., Binsaif A., Boukrami E. Business model innovation: A review and research agenda // *New England Journal of Entrepreneurship*. 2019. Vol. 22, no. 2. P. 89–108. DOI: 10.1108/NEJE-06-2019-0030.

10. Liu Y., Lee Y., Chen Y. The impact of digital capabilities and dynamic capabilities on business model innovation // *Humanities and Social Sciences Communications*. 2024. Vol. 11. DOI: 10.1057/s41599-024-02910-z.

11. Коляда А. А. Следующий уровень: стратегический менеджмент новой эпохи. М.: Альпина ПРО, 2023. 616 с. ISBN 978-5-206-00086-3.

12. Shannon C. E. A mathematical theory of communication // *The Bell System Technical Journal*. 1948. Vol. 27. P. 379–423, 623–656.

13. Lundberg S. M., Lee S.-I. A unified approach to interpreting model predictions // *Advances in Neural Information Processing Systems*. — Long Beach, 2017. — Vol. 30. — P. 4765–4774.

14. MacQueen J. Some methods for classification and analysis of multivariate observations // *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. — Berkeley: University of California Press, 1967. — Vol. 1. — P. 281–297.