

УДК 004.032.26

DOI: [10.26102/2310-6018/2025.49.2.021](https://doi.org/10.26102/2310-6018/2025.49.2.021)

Оценка потенциала нейросетевой модели дискретного выбора с сиамским обучением для задачи прогнозирования покупки недвижимости

Л.Г. Разумовский, Н.Е. Каренин✉, М.А. Герасимова

Группа компаний «РАМАХ», Москва, Российская Федерация

Резюме. В работе рассматривается задача воспроизведения процесса покупок объектов недвижимости, решение которой позволит проверять как существующие, так и будущие алгоритмы динамического ценообразования, строить предсказания предпочтений покупателей и формировать кривую спроса. В качестве решения предлагается использовать подход, основанный на использовании моделей дискретного выбора, которые широко представлены в экономической литературе и имеют обширный круг приложений в области изучения потребительского поведения и предпочтений на конкурентных рынках. В данной работе излагается новая модель дискретного выбора, использующая нейронную сеть для формирования полезности объекта недвижимости. Предлагается подход к обучению модели через сиамские нейронные сети. Также в статье предложена нестандартная архитектура основной нейросети, позволяющая избежать потери сходимости при ее обучении. В работе проводится симуляция процесса покупки объектов недвижимости с помощью классических моделей, основанных на логистической регрессии со случайными коэффициентами, и с помощью нейросетевой модели, а также проведено их сравнение. В результате численных экспериментов показано заметное преимущество предложенного нейросетевого подхода. С помощью пермутационного теста доказана статистическая значимость полученных результатов.

Ключевые слова: модель дискретного выбора, сиамские нейросети, процесс продаж, недвижимость, предпочтения покупателей, эконометрическое моделирование.

Для цитирования: Разумовский Л.Г., Каренин Н.Е., Герасимова М.А. Оценка потенциала нейросетевой модели дискретного выбора с сиамским обучением для задачи прогнозирования покупки недвижимости. *Моделирование, оптимизация и информационные технологии.* 2025;13(2). URL: <https://moitvvt.ru/ru/journal/pdf?id=1819> DOI: 10.26102/2310-6018/2025.49.2.021

Evaluating the potential of neural discrete choice model with siamese networks for a real estate sales forecasting problem

L.G. Razumovsky, N.E. Karenin✉, M.A. Gerasimova

"Ramax Group", Moscow, the Russian Federation

Abstract. The paper considers the problem of reproducing the process of purchasing real estate, the solution of which will allow testing both existing and future dynamic pricing algorithms, building predictions of buyers' preferences and forming a demand curve. As a solution, it is proposed to use an approach based on the use of discrete choice models, which are widely represented in the economic literature and have a wide range of applications in the field of studying consumer behavior and preferences in competitive markets. This paper presents a new discrete choice model that uses a neural network to form the utility of a real estate object. An approach to training the model through Siamese neural networks is proposed. The article also proposes a non-standard architecture of the main neural network, which allows avoiding the loss of convergence during its training. The paper simulates the process of purchasing real estate using classical models based on logistic regression with random coefficients and using a neural network model, and compares them. As a result of numerical

experiments, a noticeable advantage of the proposed neural network approach is shown. Using a permutation test, the statistical significance of the obtained results is proved.

Keywords: discrete choice model, siamese neural networks, sales process, real estate, customer preference, econometric modeling.

For citation: Razumovskiy L.G., Karenin N.E., Gerasimova M.A. Evaluating the potential of neural discrete choice model with siamese networks for a real estate sales forecasting problem. *Modeling, Optimization and Information Technology*. 2025;13(2). (In Russ.) URL: <https://moitvvt.ru/journal/pdf?id=1819> DOI: 10.26102/2310-6018/2025.49.2.021

Введение

Процесс выбора объекта недвижимости потенциальным покупателем представляет собой задачу, в которой необходимо предсказать вероятность выбрать конкретный объект из предлагаемых альтернатив. Принятие решения о покупке недвижимости зависит от цен, характеристик жилья, инфраструктуры, расположения, а также индивидуальных предпочтений. Для девелоперов понимание логики этого выбора имеет практическое значение: оно позволяет прогнозировать спрос, оптимизировать предложение и формировать персонализированные рекомендации. Это требует разработки моделей, которые способны воспроизводить процесс принятия решений потребителями, отражая их реальные предпочтения и структуру выбора среди конкурирующих вариантов.

Целью данного исследования является создание модели, воспроизводящей процесс выбора жилья среди конкурирующих комплексов недвижимости. Авторы ставят задачу построения последовательности выбранных объектов недвижимости, максимально близкой к истинному ходу продаж. Данная тема актуальна для девелоперов, поскольку построение такой модели позволит моделировать реальные рыночные условия и производить проверку, например, различных алгоритмов ценообразования, в условиях, когда цена ошибки высока и отсутствует возможность натурального тестирования.

Модели дискретного выбора представляют собой мощные инструменты для анализа потребительских предпочтений на конкурентных рынках. Среди приложений этих моделей к задачам реального мира особенно выделяются задачи определения предпочтений потребителей и исследование привлекательности товаров. Они находят применение в изучении выбора того, какой автомобиль купить, в какой университет поступить, каким видом транспорта добираться до работы. С помощью моделей дискретного выбора фирмы принимают различные стратегические решения, например, о том, где расположить новое торговое помещение или какие опции предложить в контракте на обслуживание.

Основополагающей работой является статья Даниэла Макфаддена [1], положившая начало современному моделированию выбора потребителей. Его методология основывается на теории максимизации случайной полезности, предполагающей, что потребитель выбирает продукт, который приносит ему наибольшую полезность. Полезность, или утилиты, – это абстрактная функция характеристик товара, показывающая насколько данный товар является привлекательным для потребителя. Полезность состоит из наблюдаемых факторов – таких как цена, характеристики продукта и ненаблюдаемых факторов – индивидуальных предпочтений, вкусов и случайных ошибок. В работе Д. Макфаддена было предложено рассматривать линейную функцию от наблюдаемых факторов полезности с добавлением случайного шума, а для обучения модели используется логистическая регрессия с кросс-энтропийной функцией потерь.

Данный подход позволяет интерпретировать влияние каждого фактора на формирование полезности товара. Однако классические logit-модели имеют ряд ограничений, снижающих их точность в реальных сценариях. Например, свойство независимости от нерелевантных альтернатив означает, что отношение вероятностей выбора между любыми двумя объектами не зависит от наличия других объектов в множестве альтернатив. В реальности это допущение неверно: когда часть объектов распродается, распределение спроса среди оставшихся квартир не является пропорциональным. Например, если в продаже были две квартиры с панорамными окнами и одна была куплена, классическая модель предположит, что вероятность выбора оставшейся квартиры изменится пропорционально. Однако в реальности покупатели могут переключиться на совершенно другой класс квартир, например, более просторные варианты без панорамных окон, что нарушает предположение независимости от нерелевантных альтернатив. Также линейная функция не способна с достаточной точностью аппроксимировать достаточно сложные зависимости полезности от исходных данных.

С ростом вычислительных возможностей стало доступным применение нейросетей для построения более сложных моделей дискретного выбора. Нейросети могут обучаться более сложным нелинейным взаимосвязям между переменными и лучше учитывать особенности данных. В отличие от logit-моделей, нейросети способны учитывать сложные нелинейные зависимости. Например, предпочтение покупателя может зависеть не просто от цены, а от комбинации факторов: дешевые квартиры, расположенные далеко от метро, менее привлекательны, чем более дорогие, но расположенные близко. Классические модели плохо отражают такие взаимосвязи, так как предполагают аддитивность характеристик.

Существует обширный круг работ, исследующих нейросетевой подход к моделям дискретного выбора, среди которых статьи Сифрингера и др. [2], Дженга и др. [3], Хадж-Йахиа и др. [4], Ауад и Дезир [5]. Помимо академических исследований, нейросетевые модели дискретного выбора находят применение в бизнесе. Например, в онлайн-рекламе нейросети применяются для прогнозирования кликов по объявлениям [6]. В ритейле используются гибридные модели для персонализированных рекомендаций [7].

Для обучения нейросетей необходим большой объем качественных данных. Несмотря на наличие крупных открытых датасетов по недвижимости, они, как правило, фиксируют лишь факты совершенных сделок, но не содержат данных об очередности продаж. В свою очередь, коммерческие данные девелоперов обычно охватывают только их собственные объекты, практически никогда не включая сведения о конкурирующих предложениях. Кроме того, поскольку покупка жилья – разовое событие, отсутствует возможность наблюдения за повторными решениями одного и того же покупателя, как это возможно, например, в ритейле. В отличие от ритейла количество объектов недвижимости фиксировано и не пополняется в ходе продаж, то есть в разные моменты времени доступно разное количество объектов. Эти факторы существенно ограничивают применение стандартных методов обучения нейросетей и требуют альтернативных подходов, пригодных к работе с малыми выборками.

В связи с этим в настоящей статье разработана нейросетевая модель дискретного выбора. Ее входной слой состоит из характеристик объекта недвижимости. На выходе нейросети для каждого экспонируемого объекта недвижимости формируется его полезность. С помощью функции softmax вычисляется вероятность выбора покупателем конкретного объекта недвижимости. Это приближает процесс принятия решений к модели Mixed-Logit, но также позволяет учесть сложные нелинейные взаимосвязи между характеристиками объектов.

Данная модель отличается от существующих специальной архитектурой нейронной сети, а также подходом к обучению, основанном на сиамских нейронных сетях, которые эффективно работают в условиях ограниченного объема данных. В отличие от стандартных нейросетевых архитектур, таких как полносвязные многослойные сети или трансформеры, сиамские нейросети обучаются не на отдельных объектах, а на упорядоченных наборах объектов, например, парах. Это делает их особенно эффективными в условиях ограниченного объема данных, поскольку каждая пара объектов создает дополнительную информацию о предпочтениях. Такой подход позволяет моделировать не только абсолютную, но и относительную полезность объектов, что лучше соответствует экономической природе процесса выбора.

В литературе широко представлено применение сиамских нейросетей в задачах, связанных с изображениями [8–10]. В подобных работах сиамские сети применяются непосредственно для обучения базовой сверточной нейросети, предназначенной для обработки изображения. Подход с применением сиамских нейронных сетей можно также использовать в задачах со схожей структурой, где присутствует выбор между альтернативными вариантами. Данный подход может быть применен к моделированию процесса покупки недвижимости, когда потенциальные покупатели выбирают между конкурирующими застройщиками. Однако, к текущему моменту в литературе не рассматривалось применение сиамских нейросетей к обучению моделей дискретного выбора, а также их применение в сфере недвижимости.

В разделе «Классические модели дискретного выбора и их применение к сфере недвижимости» мы описываем базовую модель дискретного выбора, основанную на логистической регрессии, а также приводим ее вероятностную модификацию, которую затем используем в статье для проведения экспериментов. В этом разделе мы приводим признанную в научном сообществе методологию дискретного выбора с линейной функцией полезности. В разделе «Нейросетевая модель» приводится нейросетевая модель определения утилиты, описывается архитектура и предлагается подход к обучению на основе архитектуры сиамских сетей. Также приводится обоснование выбора именно такой структуры сети и именно такого способа обучения. В разделе «Данные» описываются данные, используемые для обучения нейросетевой модели, а в разделе «Обучение и валидация» обучается модель, приводятся метрики и проверяется статистическая значимость полученных результатов, а также исследуется реакция модели на выбросы и изменения в данных. В разделе «Последовательности покупок и сравнение метрик похожести» проводится численный эксперимент – на основе обученных классических и нейросетевой моделей строятся последовательности покупок квартир, предлагаются метрики для оценки качества построенных последовательностей и приводится сравнение и интерпретация результатов работы моделей.

Основной вклад авторов статьи заключается в построении нейросетевой модели дискретного выбора и применении ее к задаче воспроизведения хода продаж. Предложенная архитектура нейронной сети приспособлена к данным из сферы недвижимости и может быть использована для работы с небольшими датасетами, что достигается за счет использования сиамских нейронных при обучении.

Классические модели дискретного выбора и их применение к сфере недвижимости

В данном разделе опишем классические модели выбора объекта недвижимости потребителем, а также их вероятностные модификации, которые будут использованы нами для проведения численных экспериментов.

Пусть имеется множество доступных объектов недвижимости $S = \{x_1, \dots, x_N\}$, каждый из которых описывается набором характеристик $x_i = \{x_{i1}, \dots, x_{iM}\}$. Задача

состоит в том, чтобы смоделировать процесс принятия решения потребителем. Традиционно это делается следующим образом. Предполагается, что потребитель в момент принятия решения присваивает каждому объекту x_i , $i = 1, \dots, N$ величину $U_i = U(x_i)$, определяющую предпочтения покупателя или привлекательность для покупателя конкретного объекта. Функция U называется функцией полезности, а ее значение на конкретном объекте – полезностью или утилитой.

В классической модели дискретного выбора функция полезности представляется в виде линейной комбинации характеристик объекта по формуле:

$$U_i = U(x_i) = a_0 + \sum_{j=1}^M a_j x_{ij}, \quad i = 1, \dots, N. \quad (1)$$

Такое представление позволяет получить аналитически удобную модель, но имеет ограничение: оно предполагает аддитивную независимость эффектов характеристик, что не всегда соответствует реальному поведению потребителей. Например, цена может зависеть от расположения, а размер квартиры – от числа комнат.

Выбранная альтернатива x^* в данной модели может рассчитываться как та, у которой утилита принимает наибольшее значение, т. е.

$$x^* = \operatorname{argmax}_{x_i \in S} U(x_i). \quad (2)$$

Возможен и другой подход, который основан на вычислении вероятностей выбора той или иной альтернативы, исходя из полученного дискретного распределения вероятностей выбора. Идея такого метода заключается в том, чтобы приблизить модель к реальности за счет случайности, которая присуща человеку при выборе между несколькими альтернативами в силу различного рода обстоятельств, которые мы не можем полностью учесть. При таком подходе выбранная альтернатива x^* представляет собой выбор из случайного распределения вероятностей $\{p_1, p_2, \dots, p_N\}$, где вероятность p_i выбора объекта x_i можно записать в виде

$$p_i = \Pr(x_i) = \operatorname{softmax}(U(x_i), \dots, U(x_N)) = \frac{\exp(U(x_i))}{\sum_{k=1}^N \exp(U(x_k))}, \quad i = 1, \dots, N. \quad (3)$$

Для того, чтобы оценить коэффициенты (веса) в этой модели, используется логистическая регрессия. Если выбор между альтернативами происходит по формуле (2), то такую модель дискретного выбора называют Max-Logit, а если по формуле (3), – Mixed-Logit.

В реальности важность характеристик объектов недвижимости может варьироваться от человека к человеку. Например, одному покупателю важнее расположение, а другому – площадь квартиры. Поэтому уместно [11] вместо детерминированных коэффициентов a_j в модели (1) рассматривать случайные коэффициенты ξ_j из некоторого фиксированного вероятностного распределения $\mathcal{P}(\theta)$:

$$U_i = U(x_i) = \xi_0 + \sum_{j=1}^M \xi_j x_{ij}, \quad i = 1, \dots, N. \quad (4)$$

В качестве вероятностного распределения $\mathcal{P}(\theta)$ можно рассматривать нормальное или логнормальное распределения с математическими ожиданиями μ_j и дисперсиями σ_j . Обучение таких моделей состоит в оценке параметров распределения $\mathcal{P}(\theta)$ также с помощью логистической регрессии. В данной работе для демонстрации этого способа решения основной задачи будут рассматриваться только эти усовершенствованные версии базовой модели дискретного выбора, которые также будут называться Max-Logit и Mixed-Logit.

Нейросетевая модель

В данном разделе опишем новую модель дискретного выбора, которая, в отличие от Max-Logit и Mixed-Logit, формирует полезность с помощью некоторой нейронной сети NN (называемой в работе *базовой*). Функция NN принимает на вход объект x_i и возвращает его предсказанную полезность U_i . Таким образом, схема выбора объекта недвижимости имеет следующий вид.

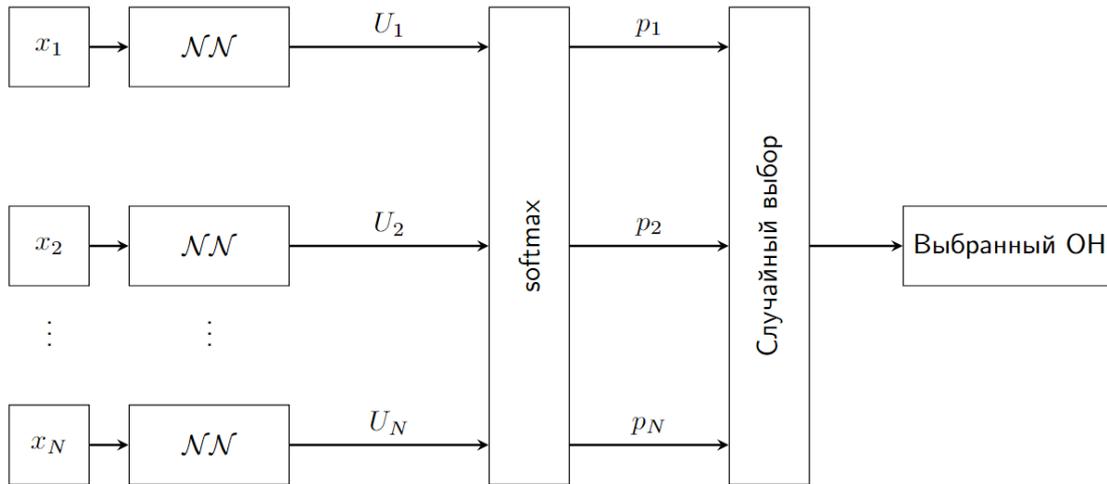


Рисунок 1 – Процесс выбора объекта недвижимости с помощью нейросети
 Figure 1 – Choosing of real estate objects using neural network

Каждый объект недвижимости x_i , $i = 1, \dots, N$ подается на вход базовой нейросети NN . На выходе формируется значение полезности U_i для i -го объекта. Далее, производится softmax-преобразование над вектором полезностей, получаем вероятность $p_i = \text{softmax}(U_1, \dots, U_N)_i$ покупки объекта x_i . Затем, на основе полученного распределения вероятностей осуществляется выбор объекта недвижимости. Таким образом, предлагаемая модель характеризует поток покупателей в целом. Выход модели случайный, это означает, что осуществляющий свой выбор покупатель сначала оценивает все доступные варианты и затем с той или иной вероятностью предпочтет купить одну из рассматриваемых квартир. Другими словами, случайный выход модели описывает как случайность выбора предпочтений покупателя, так и случайность выбора конкретного объекта недвижимости этим покупателем.

Одним из важнейших вопросов в построении данной модели является алгоритм обучения базовой нейросети. Сложность связана с тем, что в каждом процессе выбора участвует большое и, главное, разное для разных покупателей, число объектов. Также, поскольку формирование полезности объекта недвижимости напрямую связано с моделированием процесса принятия решения человеком, то предлагается использовать для обучения сиамские нейросети.

Сиамская нейронная сеть состоит из двух или более подсетей с одинаковой структурой и набором весовых коэффициентов. Данный подход может быть использован для тренировки базовой нейросети методом попарного сравнения. При этом, у каждой подсети всегда ровно один входящий вектор, что дает возможность при обучении использовать стандартные библиотеки. Схема обучения базовой сети с помощью этой парадигмы представлена на схеме:

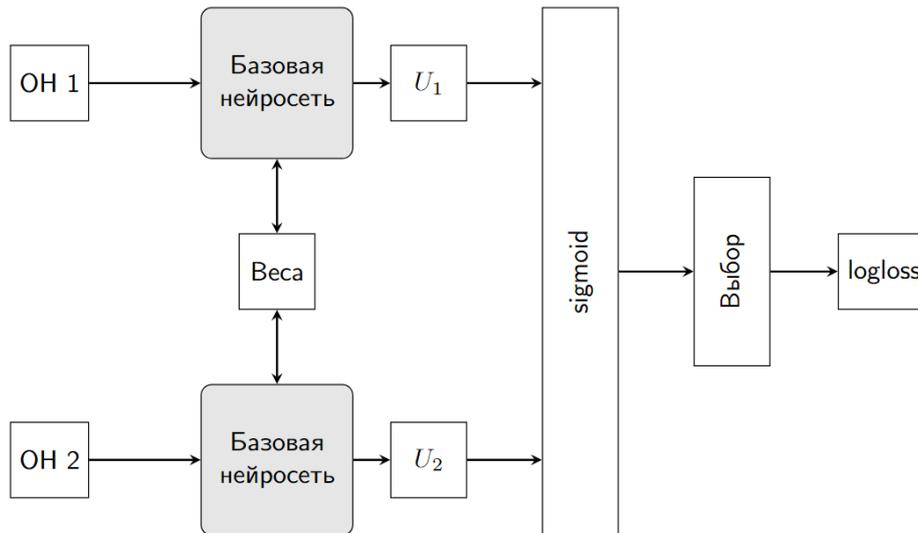


Рисунок 2 – Обучение базовой нейросети
 Figure 2 – Base network training

Одной из особенностей использования сиамских нейронных сетей при обучении данной модели, а именно применения softmax-преобразования для формирования распределения вероятностей выбора, является тот факт, что при добавлении произвольной положительной константы к утилите, вероятность выбора конкретного объекта недвижимости не изменится. При построении и обучении модели было замечено, что этим свойством можно воспользоваться для ускорения сходимости обучения. Также было обнаружено, что если вес при первом нейроне на предвыходном слое нейронной сети равен 1 и не меняется в ходе обучения, то скорость сходимости выше. В связи с этим была разработана нестандартная архитектура *NN*, в которой последний слой сети не имеет смещений, а один из весовых коэффициентов равен единице. В результате удалось добиться быстрой сходимости при обучении.

Опишем подробнее архитектуру базовой нейронной сети.

1. Входной слой: $x^{(0)} \in S = \{x_1, \dots, x_N\}$.
2. Первый обучаемый слой: $x_j^{(1)} = \sum_{i=1}^6 w_{ij}^{(1)} x_i^{(0)} + b_j, j = 1, \dots, 64$.
3. Активационный слой: $x_j^{(2)} = ReLu(x_j^{(1)}), j = 1, \dots, 64$.
4. Необучаемый слой: $x^{(3)} = x_1^{(2)}$.
5. Второй обучаемый слой: $x^{(4)} = \sum_{j=2}^{64} w_j^{(2)} x_j^{(2)}$.
6. Выход: $y = x^{(3)} + x^{(4)}$.

Выбор числа нейронов в первом и втором обучаемых слоях обусловлен следующими соображениями. Как известно [12], нейронные сети с прямой связью способны аппроксимировать зависимости произвольной сложности, если на скрытых слоях достаточно нейронов. Также слишком большое их число может привести к переобучению. Поскольку количество объектов недвижимости, предлагаемых покупателю, а также число характеристик жилья невелико, то предлагается использовать минимальное число нейронов (64 для первого и 63 для второго), необходимое для того, чтобы нейросеть обучилась сложным нелинейным зависимостям в данных, но не слишком большое, чтобы избежать переобучения.

Выходом базовой нейронной сети по смыслу является утилита объекта недвижимости. Для обучения *NN* предлагается использовать подход через сиамские нейросети. Сиамская нейронная сеть *SNN* состоит из двух идентичных подсетей, выходы

которых объединяются функцией σ , формирующей вероятность выбора одной из двух альтернатив. Формально, сиамская сеть представляет собой функцию

$$SNN(x_i, x_j, w) = \sigma(|NN(x_i, w) - NN(x_j, w)|)$$

от пары объектов (x_i, x_j) и набора весовых коэффициентов $w = (w_{ij}^{(1)}, w_k^{(2)})$.

В сиамских нейросетях разница выходов двух подсетей используется для оценки вероятности предпочтения одного объекта перед другим. Применение модуля $|U_i - U_j|$ позволяет учесть только величину различия в полезности, игнорируя направление. Функция $\sigma(\cdot)$ превращает эту разницу в вероятность предпочтения объекта x_i перед x_j . Каждой паре (x_i, x_j) поставлена в соответствие метка $y_{ij} \in \{0,1\}$: $y_{ij} = 0$, если $U_i > U_j$, и $y_{ij} = 1$, если $U_i < U_j$.

Для обучения сети используется стандартная кросс-энтропийная функция потерь, которая измеряет расхождение между предсказанными вероятностями и фактическими предпочтениями из обучающей выборки:

$$\logloss = -\frac{1}{pairs} \sum [y_{ij} \log Pr(x_i, x_j) + (1 - y_{ij}) \log (1 - Pr(x_i, x_j))], \quad (5)$$

где $Pr(x_i, x_j) = SNN(x_i, x_j, w)$, *pairs* – число пар.

Данные

В этом и следующем разделах описаны данные и обучение предлагаемой модели для последующего численного эксперимента по моделированию процесса покупки объектов недвижимости. Задача эксперимента – продемонстрировать работу нейросетевой модели, а также сравнить результаты ее работы с Max-Logit и Mixed-Logit (точнее, с их вероятностными модификациями) на данных, приближенных к действительности. В связи с тем, что измерение и сравнение качества этих моделей необходимо нам только для демонстрации их минимальной жизнеспособности и оценки их потенциального применения, подготовлен искусственный набор данных, называемый витриной. Витрина состоит из комплексов недвижимости со следующим набором характеристик: стоимость одного квадратного метра, общая площадь, количество комнат, этаж, количество санузлов, расстояние до метро. Число комплексов равно трем, и каждый из них содержит по 180 квартир на 12 этажах, в том числе 90 однокомнатных, 60 двухкомнатных и 30 трехкомнатных. Смоделирован поток покупателей в количестве $N = 3 \times 180 = 540$ – столько же, сколько всего квартир.

В данном разделе рассматривается упрощенная симуляция рынка недвижимости, включающая три жилых комплекса. Такое разбиение выбрано с целью моделирования реалистичной, но управляемой среды, где существуют конкурирующие застройщики с объектами недвижимости с разными характеристиками. Число квартир (180 на комплекс) выбрано таким образом, чтобы обеспечить достаточную вариативность данных для обучения модели, но при этом сохранить вычислительную эффективность.

Характеристики квартир были выбраны таким образом, чтобы они в целом отражали действительность на рынке жилья. Площадь одной комнаты в квартире была выбрана следующим образом: для однокомнатных – $35,0 \pm 5,0$ м², двухкомнатных – $30,0 \pm 5,0$ м² и трехкомнатных – $27,5 \pm 7,5$ м². Параметры площадей выбраны на основе средних показателей у застройщиков в крупных городах. Разброс значений отражает вариативность предложения.

Расстояние до метро было выбрано таким образом, чтобы удаленность каждого комплекса была различной – это значения 1, 2 и 4 расстояния в у. е. Цена за 1 м² была

назначена в зависимости от удаленности от метро, а также в зависимости от этажа – чем выше расположена квартира в комплексе, тем она дороже. Наконец, для однокомнатных и двухкомнатных квартир был взят один санузел, а для трехкомнатных – два.

Сгенерирована последовательность продаж объектов недвижимости, то есть каждому объекту сопоставлена очередность его покупки. Для корректного моделирования процесса продаж недвижимости требуется учитывать не только характеристики самих объектов, но и динамику их выбора покупателями. Однако доступность таких данных крайне ограничена. Открытые источники фиксируют только окончательные сделки, не предоставляя информации о том, какие альтернативы были доступны покупателю на момент выбора. Коммерческие данные девелоперов, даже если они доступны, обычно охватывают лишь объекты одной компании и не содержат информации о предложениях конкурентов. Более того, покупка недвижимости является разовым событием, что исключает возможность накопления статистики повторных решений одних и тех же покупателей, как в ритейле. В совокупности это приводит к дефициту данных, достаточных для обучения нейросетей традиционными методами.

В связи с этим в данной работе применяется сиамский подход с попарным сравнением. Такой метод увеличивает объем доступной информации за счет представления данных в виде пар сравнений объектов, позволяя эффективно обучать модель даже в условиях ограниченности выборки. При таком подходе базовая нейронная сеть учится предсказывать наиболее привлекательный объект из пары. В свою очередь, обученная нейросеть за счет правильного выбора архитектуры может предсказывать наиболее привлекательный объект из набора, состоящего из произвольного количества объектов недвижимости.

Для обучения базовой нейросети на основе витрины сформирован набор данных, состоящий из упорядоченных пар (x_i, x_j) следующим образом: если объект x_i был продан позже объекта x_j , то в датасет заносится только пара (x_i, x_j) . Пары с повторяющимися компонентами не заносятся. Таким образом, каждому объекту ставится в соответствие метка $y_{ij} = 1$. Отсутствие метки 0 естественно, так как заранее зафиксирован порядок в парах объектов. Обучение сводится к нахождению весовых коэффициентов, при которых выбор объекта с бóльшей утилитой наиболее вероятен. Таким образом, это соотносится с исходной задачей обучить модель правильно определять наиболее предпочтительную недвижимость. Стоит подчеркнуть, что на стадии обучения базовой нейросети нам нужно оценивать только правильность определения объекта внутри пары по его утилите на основании исторической очередности продаж, а не качество ранжирования. То есть показателем качества модели должна служить метрика, характеризующая количество пар, в которых наиболее привлекательный объект недвижимости был выбран правильно. Поскольку в нашем случае все пары, отнесенные к классу 1, будут классифицированы корректно (у нас отсутствуют метки 0), можно ограничиться ассигасу, которая покажет долю верных предсказаний, то есть положительно классифицированных пар. Чем выше значение ассигасу, тем точнее модель.

Полученный датасет разбивается на тренировочную, валидационную и тестовую выборки так, как это представлено на диаграмме (Рисунок 3):



Рисунок 3 – Разбиение датасета, состоящего из пар
Figure 3 – Splitting the pairs dataset

Обучение и валидация

Далее, с помощью библиотеки TensorFlow обучим нашу модель на тренировочной выборке. Гиперпараметры модели были подобраны на основе предварительных экспериментов с использованием кросс-валидации.

- Число эпох (10): выбрано на основе анализа кривых обучения. После 10 эпох наблюдается стабилизация метрик без значительного переобучения.
- Размер батча (32): баланс между скоростью обучения и стабильностью градиента.
- Скорость обучения (0,001): обеспечивала оптимальную сходимость Adam-оптимизатора, предотвращая слишком резкие скачки в градиентах.

На Рисунках 4 и 5 представлены кривые обучения модели. График accuracy демонстрирует монотонный рост, что свидетельствует о стабильном обучении. Функция потерь снижается и стабилизируется, что говорит о хорошем балансе между bias и variance. Отсутствие резкого роста ошибки на валидационной выборке указывает на отсутствие явного переобучения.

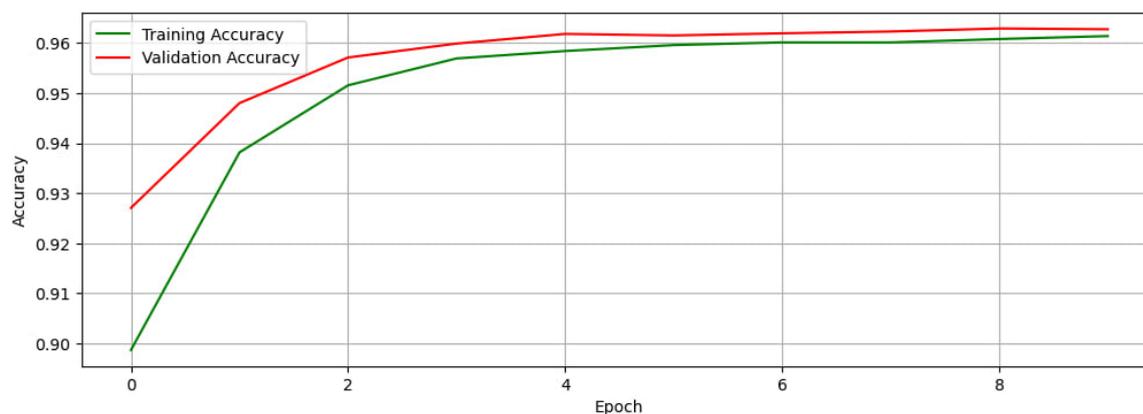


Рисунок 4 – Кривая обучения: точность
Figure 4 – Learning curve: accuracy

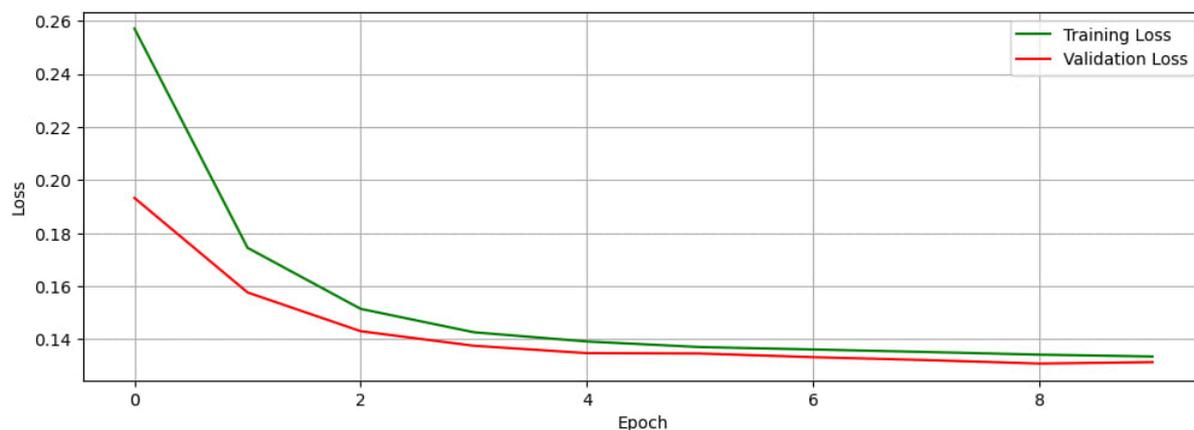


Рисунок 5 – Кривая обучения: функция потерь
Figure 5 – Learning curve: loss

На тестовой выборке метрика accuracy = 0,962, что говорит о высоком качестве предсказания, такая модель редко пропускает истинные предпочтения покупателей.

Последовательности покупок и сравнение метрик похожести

В данном разделе мы применим обученную нейросетевую модель к построению последовательности покупок объектов недвижимости из витрины. Также проведем сравнение этой модели с Max-Logit и Mixed-Logit.

В рамках той или иной модели в процессе выбора утилитарный покупатель рассчитывает полезность каждого объекта, после чего либо выбирает тот объект, у которого она наибольшая, либо наиболее вероятный вариант.

На основе данного алгоритма выбора можно построить последовательность продаж всех объектов недвижимости из витрины, после чего представляется возможным изучать предпочтения покупателей, делать какие-либо выводы о привлекательности конкретного типа недвижимости и т. д. Построить последовательность выбора представляется естественным поэтапно, то есть для каждого момента выбора $i = 1, \dots, N$ покупатель выбирает из $N - i + 1$ объектов недвижимости из витрины, которая с каждым шагом становится меньше на один выбранный объект недвижимости. В каждой итерации покупатель выбирает объект недвижимости из доступного множества альтернатив. Выбор осуществляется согласно одной из рассмотренных моделей:

1. Max-Logit: выбирается объект с наибольшей полезностью U_i .
2. Mixed-Logit: выбор осуществляется случайно, с вероятностями, заданными softmax от полезностей U_i .
3. Нейросетевая модель: аналогично Mixed-Logit, но вместо линейной функции полезности используются выходы нейросети.

После выбора объект удаляется из множества доступных, и процесс повторяется до тех пор, пока все объекты не будут распределены.

В литературе изучалось также применение трансформеров [13] и бустинговых моделей [14] к моделированию потребительского выбора. В контексте нашей задачи бустинг-модели не подходят по ряду причин: склонность к переобучению на малых выборках, что приводит к запоминанию случайного шума в распределении признаков, а не выявлению закономерностей; излишняя сложность, в то время как простые модели имеют лучшую обобщающую способность и меньшую вычислительную нагрузку; учитывая низкую размерность задачи, использование бустинга только увеличивает разброс, что делает его заметно менее эффективным. В свою очередь, для трансформеров

характерна работа с эмбедингами, включающими в себя десятки тысяч характеристик, что не представляется возможным при работе с данными о продажах жилья. Поэтому их также было решено не включать в рассмотрение.

Для того, чтобы сравнить классические модели с нейросетевой моделью, создадим эталонную последовательность $(\pi(1), \dots, \pi(N))$, где π – перестановка набора $(1, \dots, N)$. Последовательность, сгенерированную одной из моделей, обозначим $(\pi^{-1}(1), \dots, \pi^{-1}(N))$, π^{-1} – также некоторая перестановка, вообще говоря, несовпадающая с π . Для оценки соответствия предсказанных последовательностей истинной используем ранговые коэффициенты корреляции:

1. Коэффициент Кендалла τ измеряет долю пар объектов, расположенных в правильном порядке относительно эталонной последовательности. Чем выше τ , тем больше совпадений в порядке предпочтений.

2. Коэффициент Спирмена ρ измеряет степень монотонной зависимости между истинным и предсказанным порядком. Он чувствителен к глобальным искажениям ранжирования.

Оба коэффициента наиболее естественно интерпретируются в задачах ранжирования, так как учитывают не абсолютные значения полезностей, а относительное расположение объектов в последовательности. Также выбор данных коэффициентов обоснован еще и тем, что в отличие от других метрик, таких как NDCG или Precision@k, коэффициенты Кендалла и Спирмена глобально измеряют похожесть структур порядка в последовательностях, а не только в нескольких первых ее членах.

Поскольку непосредственно выбор объекта недвижимости происходит случайно, то имеет смысл находить значения τ и ρ не для единственной предсказанной последовательности, а для нескольких последовательностей, и затем рассчитать их среднее значение. В результате эксперимента получили следующие значения коэффициентов корреляции:

Таблица 1 – Ранговая корреляция
Table 1 – Rank correlation

Модель	Коэффициент Кендалла	Коэффициент Спирмена
Max-Logit	0,58474638	0,76423965
Mixed-Logit	0,58472480	0,76426377
Нейросеть	0,85537333	0,97109835

В Таблице 1 приведены значения коэффициентов корреляции для различных моделей. Логит-модели (Max-Logit, Mixed-Logit) показывают примерно одинаковые результаты ($\tau \approx 0,58$, $\rho \approx 0,76$), что указывает на их схожесть в предсказаниях. Данные значения являются довольно высокими, несмотря на значение коэффициента Кендалла, который очень чувствителен к отклонению от истинной последовательности. Нейросетевой подход значительно превосходит логит-модели ($\tau \approx 0,86$, $\rho \approx 0,97$), что свидетельствует о более точном восстановлении порядка покупок.

Для проверки статистической значимости превосходства нейросетевой модели проведен пермутационный тест (10000 перестановок). Полученное р-значение составило 0,003, что ниже порогового уровня $p < 0,05$. Это подтверждает, что улучшение качества предсказаний не случайно, а обусловлено структурой нейросетевой модели.

Заключение

Выводы. Разработана модель, воспроизводящая процесс выбора жилья среди конкурирующих комплексов недвижимости. Эта модель основана на специальной

архитектуре базовой нейросети, учитывающей особенности задачи и обеспечивающей устойчивую сходимость при обучении. Выполнено построение последовательности выбранных объектов недвижимости, моделирующей реальную очередность продаж, с помощью предложенной в данной работе нейросетевой модели дискретного выбора с сиамским подходом к обучению. Проведено сравнительное исследование с классическими моделями и показано, что нейросетевая модель превосходит их по метрикам ранговой корреляции.

Дальнейшее развитие.

– Графовые нейросети (GNN): они позволяют учитывать связи между объектами (например, конкурирующие предложения в одном районе). Это может улучшить качество предсказаний за счет моделирования локальных эффектов.

– Временные модели: применение рекуррентных сетей (LSTM, Transformer) для учета эволюции спроса во времени. Они могут быть использованы в качестве дополнительного модуля при генерации процесса покупки объектов недвижимости, что значительно улучшит качество воспроизведения последовательности продаж. Такое дополнение сделает модель более приспособленной к решению реальных задач бизнеса.

– Байесовские методы: изучение неопределенности предсказаний с помощью вероятностных моделей позволит увеличить интерпретируемость результатов и их надежность.

Таким образом, предложенная в работе нейросетевая модель дискретного выбора представляет собой новый метод анализа потребительских предпочтений, расширяющий возможности традиционного эконометрического моделирования. Полученные результаты демонстрируют преимущества предлагаемого подхода, а дальнейшее развитие методологии открывает возможности для более точного моделирования поведения потребителей, что может быть использовано застройщиками для симуляции процесса покупки комплекса недвижимости и с минимальными затратами тестировать в них различные алгоритмы оптимизации выручки.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. McFadden D. Economic Choices. *American Economic Review*. 2001;91(3):351–378. <https://doi.org/10.1257/aer.91.3.351>
2. Sifringer B., Lurkin V., Alahi A. Enhancing Discrete Choice Models with Neural Networks. In: *18th Swiss Transport Research Conference (STRC 2018), 16–18 May 2018, Monte Verità, Switzerland*. 2018. pp. 1–13.
3. Jeng J.-M., Fesenmaier D.R. A Neural Network Approach to Discrete Choice Modeling. *Journal of Travel & Tourism Marketing*. 1996;5(1-2):119–144. https://doi.org/10.1300/J073v05n01_08
4. Haj-Yahia Sh., Mansour O., Toledo T. Incorporating Domain Knowledge in Deep Neural Networks for Discrete Choice Models. arXiv. URL: <https://arxiv.org/abs/2306.00016> [Accessed 20th December 2024].
5. Aouad A., Désir A. Representing Random Utility Choice Models with Neural Networks. arXiv. URL: <https://arxiv.org/abs/2207.12877> [Accessed 20th December 2024].
6. Yang Ya., Zhai P. Click-Through Rate Prediction in Online Advertising: A Literature Review. *Information Processing & Management*. 2022;59(2). <https://doi.org/10.2139/ssrn.4036054>
7. Guo Ya., Wang M., Li X. An Interactive Personalized Recommendation System Using the Hybrid Algorithm Model. *Symmetry*. 2017;9(10). <https://doi.org/10.3390/sym9100216>
8. Craparotta G., Thomassey S., Biolatti A. A Siamese Neural Network Application for Sales Forecasting of New Fashion Products Using Heterogeneous Data. *International*

- Journal of Computational Intelligence Systems*. 2019;12:1537–1546. <https://doi.org/10.2991/ijcis.d.191122.002>
9. Van Cranenburgh S., Garrido-Valenzuela F. Computer vision-enriched discrete choice models, with an application to residential location choice. arXiv. URL: <https://arxiv.org/abs/2308.08276> [Accessed 20th December 2024].
 10. Ruan G., Kirschen D.S., Zhong H., Xia Q., Kang C. Estimating Demand Flexibility Using Siamese LSTM Neural Networks. *IEEE Transactions on Power Systems*. 2021;37(3):2360–2370. <https://doi.org/10.48550/arXiv.2109.01258>
 11. Train K.E. *Discrete Choice Methods with Simulation*. Cambridge; New York: Cambridge University Press; 2009. 388 p. <https://doi.org/10.1017/CBO9780511805271>
 12. Hornik K. Approximation capabilities of multilayer feedforward networks. *Neural Networks*. 1991;4(2):251–257. [https://doi.org/10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T)
 13. Dong G., Kweon Y., Park B.B., Boukhechba M. Utility-Based Route Choice Behavior Modeling Using Deep Sequential Models. *Journal of Big Data Analytics in Transportation*. 2022;4(2-3):119–133. <https://doi.org/10.1007/s42421-022-00058-3>
 14. Wang F., Ross C.L. Machine Learning Travel Mode Choices: Comparing the Performance of an Extreme Gradient Boosting Model with a Multinomial Logit Model. *Transportation Research Record: Journal of the Transportation Research Board*. 2018;2672(47). <https://doi.org/10.1177/0361198118773556>

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Разумовский Лев Григорьевич, кандидат технических наук, президент ГК, Группа компаний «Рамак», Москва, Российская Федерация.

e-mail: lev.razumovskiy@ramax.com
ORCID: [0009-0009-7857-3101](https://orcid.org/0009-0009-7857-3101)

Lev G. Razumovskiy, Candidate of Engineering Sciences, President, "Ramax Group", Moscow, the Russian Federation.

Каренин Николай Евгеньевич, математик-программист, Группа компаний «Рамак», Москва, Российская Федерация.

e-mail: nikolay.karenin@ramax.com
ORCID: [0009-0000-1317-2512](https://orcid.org/0009-0000-1317-2512)

Nikolay E. Karenin, programming mathematician, "Ramax Group", Moscow, the Russian Federation.

Герасимова Мария Алексеевна, кандидат физико-математических наук, руководитель Центра перспективных решений, Группа компаний «Рамак», Москва, Российская Федерация.

e-mail: mariya.gerasimova@ramax.com

Mariya A. Gerasimova, Candidate of Physico-Mathematical Sciences, Head of R&D Center, "Ramax Group", Moscow, the Russian Federation.

Статья поступила в редакцию 06.02.2025; одобрена после рецензирования 19.03.2025; принята к публикации 08.05.2025.

The article was submitted 06.02.2025; approved after reviewing 19.03.2025; accepted for publication 08.05.2025.