

DOI: 10.17323/2587-814X.2024.4.7.24

Нейросетевые технологии в управлении цепями поставок: методика выбора покупателя*

В.А. Назаркина 

E-mail: valeria71@bk.ru

В.Ю. Щеколдин 

E-mail: schekoldin@corp.nstu.ru

Новосибирский государственный технический университет, Новосибирск, Россия

Аннотация

Эффективность управления цепями поставок зависит, в том числе, от выбора и слаженного взаимодействия с покупателями продукции. Статья посвящена разработке методики выбора покупателя на региональном оптово-розничном рынке топлива. Методологической базой исследования является теория многомерного статистического анализа и нейронных сетей. Основным инструментом для разработки методики послужили нейросетевые технологии, при помощи которых можно корректно оценить границы диапазона значений показателей, характеризующих потребителей и отражающих их историю покупательского поведения, для выбора потенциальных клиентов и возможности дальнейшего сотрудничества с уже существующими. Информационную базу работы составили данные о покупателях продукции, данные электронного справочника 2ГИС, результаты проведенного первичного статистического анализа и построения прогнозов, сделанных на основе нейросетей различной топологии. Предложена авторская методика выбора покупателя, имеющая потенциал развития и использования для решения ряда других управленческих задач. В рамках апробации была определена наиболее подходящая конфигурация нейросети, и оценены стандартные значения входных барьеров для выбора покупателя. Апробация разработанной методики была проведена на примере компании, функционирующей на оптово-розничном рынке топлива в г. Новосибирске и Новосибирской области. В процессе верификации модели нейронной сети было проведено сопоставление качества классификации клиентов на основе моделей логистической регрессии, дерева принятия решений и случайного леса, и обнаружено, что нейросетевой подход обеспечивает наилучшие результаты оценивания степени пригодности клиентов. Разработаны рекомендации по совершенствованию нейросетевых моделей, включающие расширение набора факторов, которые определяют характеристики потребителей, а также оптимизацию внутренней структуры нейронных сетей.

* Статья опубликована при поддержке Программы НИУ ВШЭ «Университетское партнерство»

Ключевые слова: нейронные сети, управление цепями поставок, логистика распределения, методика выбора покупателя, обучение нейросети

Цитирование: Назаркина В.А., Шеколдин В.Ю. Нейросетевые технологии в управлении цепями поставок: методика выбора покупателя // Бизнес-информатика. 2024. Т. 18. № 4. С. 7–24.
DOI: 10.17323/2587-814X.2024.4.7.24

Введение

Непредсказуемость влияния факторов внешнего окружения ставит перед менеджментом организаций задачи, связанные с риском выбора неверного управленческого решения в различных направлениях деятельности компании. Несмотря на стремление привлечь как можно большее число покупателей и заручиться их лояльностью, организация сталкивается с проблемами сложности взаимодействия, неэффективности сделок и невыполнением договорных обязательств со стороны контрагентов. В теории и на практике выработано достаточное количество методик выбора надежного поставщика ресурсов, однако вопросами выбора покупателя ученые и представители бизнеса детально не занимаются. Зачастую решение вопроса о выборе и дальнейшей работе с ними определяется методом проб и ошибок, что в итоге сказывается на эффективности управленческих решений в области управления цепями поставок в целом и на взаимодействиях с потребителями в частности.

На сегодняшний день одним из актуальных методов решения самых разнообразных экономических и управленческих задач считается применение нейросетевых технологий. С точки зрения логистики значимыми объектами для применения нейронных сетей являются процессы управления цепями поставок, поскольку с помощью этих методов их можно оценивать, моделировать и прогнозировать развитие с заранее заданной точностью.

Цель исследования – разработка методики выбора покупателей на основе применения нейросетей и ее апробация на примере организации, функционирующей на региональном оптово-розничном рынке топлива.

1. Теоретические аспекты исследования

Нейронные сети представляют собой определенный тип моделей искусственного интеллекта, базирующийся на структуре, динамике и функциях человеческого мозга. В самом общем виде нейросети состоят из множества взаимосвязанных между собой узлов (нейронов), и используются для обработки, моделирования и анализа сложных неоднородных процессов. Нейронные сети привлекают широкое внимание и интенсивно применяются в различных областях благодаря их способности не просто анализировать данные, но и самостоятельно строить логические выводы, формировать оценочные суждения и разрабатывать прогнозные решения самого разнообразного профиля.

Исторически первые работы, связанные с приложением теорий, описывающих функционирование мозга, появились в 40–50-е года прошлого века, когда появилась необходимость построения концепций искусственных нейронов для определения их потенциала при создании сложных интеллектуальных систем. Одним из первых удачных представлений нервной системы стала предложенная Ф. Розенблаттом в 1957 году модель перцептрона [1], которая оказалась способна решать некоторые задачи распознавания. Несмотря на первые успехи модели перцептрона, в ней нашлись определенные недостатки, не дающие возможность решения на ее основе определенных типов задач. В 1970–80-х годах был разработан алгоритм обратного распространения ошибки [2], который позволил обучать более сложные сети. Среди многочисленных работ, посвященных рассматриваемым вопросам, следует отметить существенный вклад в развитие теории нейронных сетей и распознавания образов советского и российского ученого А.И. Галушкина [3].

Свою лепту в развитие и распространение нейросетевых технологий внес и XXI век, можно отметить резкое увеличение разработок, связанных с так называемым глубоким обучением (deep learning). Пионерские исследования в этой области связаны с именем Дж. Хинтона, который внес значительный вклад в разработку современных алгоритмов обучения глубоких нейронных сетей [4]. Глубокое обучение включает в себя обучение нейросетей с несколькими скрытыми слоями, в том числе с динамически меняющейся структурой, наличием внешних и внутренних топологий и т.д.

Следует отметить, что наряду с нейросетевыми моделями на практике встречаются и другие подходы, позволяющие решать различные задачи классификации. К таким методам относятся, например, построение моделей логистической регрессии, деревьев принятия решений, моделей случайного леса и т.д. [5]. Отметим, что каждый из этих подходов обладает своими преимуществами и недостатками. Рассмотрим некоторые из них.

К преимуществам моделей логистической регрессии относятся [6]: высокая степень интерпретируемости результатов, большая эффективность статистических процедур, используемых для оценивания параметров модели; гибкость при решении бинарных и многофакторных задач классификации. Модели логистической регрессии обладают и рядом недостатков: зависимость получаемых результатов от структуры входных факторов; высокая чувствительность к наличию выбросов в исходных данных; частое возникновение эффекта мультиколлинеарности.

Применение моделей деревьев принятия решений обеспечивает построение легкой и удобной интерпретации получаемых решений; универсальность вычислительной схемы для любых типов данных; робастность результатов классификации, т.е. устойчивость к выбросам в исходных данных. Минусы данного подхода состоят в том, что получаемые деревья принятия решений часто имеют сложную и запутанную структуру (склонность к переобучению); возникает неустойчивость классификации к изменениям в исходных данных; часто появляются локально-оптимальные решения, которые определяются неоднородностями исходных данных или изучаемых процессов [7].

При использовании моделей случайного леса исследователь получает следующие преимущества: возможность масштабирования и удобного распараллеливания базового алгоритма в зависимости от

свойств решаемой задачи; возможность ранжирования независимых факторов, включаемых в модель; высокую эффективность классификаций для задач больших размерностей. При этом такие модели обладают следующими недостатками: сложная и неоднозначная структура модели из-за большого разнообразия вариантов усреднения; неустойчивость к флуктуациям в исходных данных; большое число эмпирически определяемых параметров алгоритма; высокие требования к характеристикам вычислительной техники (память, быстродействие и т.д.) [8, 9].

Развитие нейросетевых технологий оказало большое влияние на сферу логистической деятельности. Многие постановки задач логистики и их методы решений претерпели существенные изменения в свете теории нейросетей, перечислим некоторые из них.

В работе [10] авторы применяют идею построения многослойной нейронной сети для анализа данных о заказах, включающих количество посещений веб-сайта компании, время посещения, а также учитывающие рабочие, будние и праздничные дни. В статье [11] рассматривается применение нейросети Хопфилда для решения задачи построения динамически оптимального маршрута в телекоммуникационной сети и предлагается эвристическое правило для остановки процесса обучения нейросети, которое позволяет эффективно ограничивать время обучения.

Нейронные сети можно применять для оптимизации выполнения складских операций. В статье [12] авторы доказывают, что при управлении складом наиболее эффективно работают сети со специальными функциями активации нейронов типа Traingdx при использовании трехслойных нейросетей с топологией 6-8-1.

Еще одной сферой применения нейросетей является анализ и прогнозирование рисков в логистике. В статье [13] рассматривается задача безопасного облета препятствий пилотируемыми и беспилотными летательными аппаратами, где автор предлагает использовать многослойную сеть последовательного распространения ошибок с тремя слоями.

Следует отметить, что среди большого числа работ, посвященных решению задач логистического характера, в которых применялись бы нейросетевые технологии, практически отсутствуют те, в которых исследованию подвергались потребители товаров и услуг. В то же время изучению истории покупательского поведения и построению классификаций потребителей уделяется достаточно большое внимание

при решении задач не из области логистики [14–17], поскольку понимание структуры клиентской базы и динамики ее изменений дает в руки компании дополнительные инструменты по совершенствованию эффективности взаимодействий с клиентами.

В современной быстроменяющейся и часто плохо прогнозируемой экономической среде компаниям приходится иметь дело с ситуациями, в которых важно определять приоритеты при сотрудничестве с теми или иными клиентами. Так, на этапе заключения контрактов, бывает необходимо своевременно выявить, является ли тот или иной клиент финансово независимым, платежеспособным, не возникнут ли сложности с выполнением заказов из-за географической удаленности, возможно ли совместное использование складских мощностей и т.д. Безусловно, анализируя историю покупательского поведения, возможно определить, какие из уже заключенных и выполненных контрактов оказались для компании выгодными, какие не привели к достижению поставленных целей, а какие и вовсе оказались провальными с точки зрения получения выручки, затрат ресурсов и времени сотрудников, отразились на ухудшении имиджа компании и т.д. Эту информацию можно использовать для того, чтобы, зная

характеристики очередного потенциального клиента, заранее определить степень выгодности сотрудничества с ним. На решение этой задачи и направлено предлагаемое исследование.

2. Методика исследования

Необходимость решения задач оптимизации логистического процесса определило возможность использования технологии нейросетей в части определения границ диапазона значений для выбора потенциального покупателя, и его значимости для компании с точки зрения прибыльности и выполнения договорных обязательств. Данное исследование было структурировано на основе эмпирического материала, собранного из вторичных и первичных источников. Для решения задач исследования были использованы метод первичного статистического анализа, корреляционно-регрессионный анализ, метод обратного распространения ошибки, методы интеллектуального анализа данных [2, 6, 18, 19].

Содержание и ожидаемые результаты от применения методики выбора покупателя представлены в *таблице 1*.

Таблица 1.

Основные этапы методики выбора покупателя с помощью нейросети

№	Название этапа	Содержание	Результат
1	Описание проблемы	Анализ этапов логистического цикла, выявление проблем	Определение необходимости поиска актуальных методов получения информации о потребителях
2	Определение факторов, влияющих на процесс выбора покупателя	Формирование набора факторов, подлежащих количественной оценке специалистами компании, экспертами, заказчиками исследования	Перечень факторов, характеризующих свойства покупателей, для построения нейросети
3	Создание базы данных по актуальным покупателям организации	Определение значений выбранных факторов для оценивания покупателей	База актуальных данных о потребителях и заключенных ими сделках для обучения нейросети
4	Первичный статистический анализ	Вычисление и интерпретация статистических характеристик факторов	Предварительные выводы о свойствах наблюдаемых объектов (потребителей)
5	Определение ролей факторов	Построение простейшей нейронной сети	Ранжированный список факторов для принятия решения по выбору покупателя
6	Построение нейросети сложной структуры	Построение нейросетей различной структуры и сравнение с результатами, получаемыми на основе других моделей классификации	Выбор наилучшей конфигурации (топологии) нейросети
7	Применение «наилучшей» нейросети для идентификации перспективных клиентов	Оценивание коридоров значений факторов, при которых определяется статус покупателя	Определение стандартных значений входных барьеров для выбора покупателя

На *первом этапе* методики необходимо описать сущность проблем в части взаимодействия с потребителями компании. Анализируются этапы логистического цикла, характерного для работы с потребителями. Проблемы, возникающие в процессе распределения продукции, могут касаться вопросов информационного взаимодействия компании с покупателями, контроля качества и перевозок, их документационного обеспечения и т.д. В результате выявления проблемы руководством организации принимается решение о поиске наилучших методов выбора покупателей.

На *втором этапе* необходимо выбрать показатели, которыми руководствуется организация, принимая решение о взаимодействии с покупателем. К факторам, характерным для выбора покупателя, можно отнести: расстояние до места поставки; количество видов продукции, подлежащих одновременной реализации; мощность оборудования; рейтинг; объем продаж за определенный период времени и т.д.

На *третьем этапе* необходимо разработать и сформировать базу данных для обучения нейросети. Для всех потенциальных потребителей компании определяются значения факторов, выбранных на втором этапе. Часть факторов оценивается путем прямых измерений соответствующих показателей, а другая часть требует получения и использования вторичной информации. Кроме того, каждый из потенциальных покупателей должен быть оценен специалистом в области логистики на предмет приоритетности выбора его как реального клиента.

Четвертый этап предполагает вычисление основных статистических характеристик, определяемых по значениям факторов построенной на третьем этапе базы данных покупателей. Для обеспечения корректности статистического анализа, в частности, для определения однородных групп, внутри которых можно говорить об определенной идентичности анализируемых объектов (покупателей), данные необходимо распределять (группировать) по классам однородности, число которых k определяется, например, по формуле Стерджеса [20]

$$k = [1 + 3,32 \lg(N)], \quad (1)$$

где N – общее число наблюдений величины (объем выборки или объем базы данных);
 $\lg(.)$ – десятичный логарифм;
 $[.]$ – операция взятия целой части числа.

В результате проводится статистический анализ значений мер положения, разброса и формы, делаются предварительные выводы о свойствах потребителей и истории их покупательского поведения.

На пятом этапе осуществляется ранжирование факторов по степени влияния на принятие решения о взаимодействии с потенциальными клиентами. Для этого строится простейшая нейросеть, состоящая из одного нейрона и входных переменных, соответствующих выбранным на третьем этапе факторам. Степень и характер влияния факторов будут определяться значениями весовых коэффициентов сети. Как правило, эффективность такой сети оказывается достаточно низкой, что не позволит применять ее для прогнозирования приоритетности клиентов. Однако она позволяет проранжировать входные факторы, что дает возможность построить качественную интерпретацию процесса принятия решений по выбору клиентов.

На шестом этапе осуществляется разработка и обучение нейросетей различной топологии для обеспечения наилучшей степени прогнозируемости «нужности» клиента. При этом необходимо рассмотреть несколько вариантов сетей, отличающихся друг от друга числом скрытых слоев, их взаимосвязями между собой, а также числом нейронов на каждом слое [21]. Результатом этого этапа будет выбор такой конфигурации нейросети, которая обеспечивает наименьший уровень ошибки при прогнозировании надежности клиентов.

Для обеспечения адекватности получаемых результатов необходимо провести сопоставление качества классификации клиентов, получаемой на основе нейросетей, и классификаций, строящихся другими методами интеллектуального анализа данных. Простейшим способом сопоставления различных классификаций является применение метода таблиц сопряженности (contingency tables) [22]. Будем считать, что наилучшей будет являться та классификация, которая обеспечивает наименьшее число ошибок прогнозирования.

На седьмом этапе наилучшая из нейросетей, построенных на шестом этапе, применяется для определения коридоров значений факторов, при которых сохраняется статус текущего клиента. Это также позволит определить стандартные (базовые) значения входных факторов для упрощения процедуры выбора нового клиента.

3. Практические аспекты исследования: апробация методики

Апробация методики проводилась на основе информации, предоставленной компанией, специализирующейся на продаже сжиженного газа населению и организациям в г. Новосибирске и Новосибирской

области. Организация также осуществляет реализацию сопутствующих товаров и услуг, проводит проектные работы, оказывает сервисные услуги и организует техническое обслуживание газопроводов природного газа.

На *первом этапе* были рассмотрены операции логистического цикла. В качестве покупателей определены автогазозаправочные станции (далее – АГЗС), расположенные на территории г. Новосибирска и Новосибирской области. Изучение логистического процесса в компании позволило обнаружить необходимость регулярной проверки информации о характеристиках покупателей, так как отсутствуют четкие границы значений факторов, которые дают возможность идентифицировать покупателей со свойствами, позволяющими их считать пригодными для сотрудничества. В связи с этим возникают ошибки, причинами которых являются, например, влияние внешних факторов, ограниченность ресурсов, ошибки специалистов и др., что может приводить к неверной идентификации покупателей и, как следствие, к принятию неверных решений по работе с ними.

Руководством организации было принято решение о поиске путей оптимизации взаимодействия с покупателями в части определения стандартных значений факторов как для текущих клиентов, так и для выбора и взаимодействия с потенциальными клиентами.

На *втором этапе* был составлен пул факторов, определяющих характеристики клиентов, учитываемые при принятии решений. В этот пул вошли следующие показатели:

- ♦ расстояние от компании до клиента (км);
- ♦ ширина ассортимента (количество видов сжиженного газа, приобретаемое клиентом);
- ♦ производственная мощность (количество колонок на АГЗС);
- ♦ рейтинг АГЗС;
- ♦ суммарный объем (емкость) основных и дополнительных цистерн на АГЗС;
- ♦ средний объем продаж топлива за сутки (тыс. л).

На *третьем этапе* разработана структура базы данных, содержащая информацию о сделках компании. База данных содержала результаты деятельности компании за период сентябрь–ноябрь 2023 г.

Специалист компании, занимающий должность руководителя отдела логистики, провел оценку выгоды совершенных сделок, в результате чего все записи базы данных были промаркированы значениями бинарной переменной, определяющей «надежность» клиентов: обозначение «1» использовалось для «подходящих» клиентов, с которыми компании выгодно сотрудничать и далее, а «0» – для «неподходящих».

На *четвертом этапе* проводился первичный статистический анализ по каждому из факторов, которые характеризуют деятельность потребителей АГЗС. Вычислялись значения мер положения (среднее, медиана), мер разброса (стандартное отклонение, нижняя и верхняя границы), мер формы (коэффициенты асимметрии и эксцесса), а также экстремальные значения (минимум и максимум), результаты сведены в *таблицу 2*.

Таблица 2.

Значения статистических характеристик факторов

Статистические показатели	Факторы и их обозначения					
	Расстояние, км	Ширина ассортимента, ед.	Производ. мощность ед.	Рейтинг, усл. ед.	Емкость цистерн, тыс. л	Объем продаж, тыс. л
	x_1	x_2	x_3	x_4	x_5	x_6
Среднее	118,469	2,563	4,547	2,469	19,983	5,610
Медиана	51,500	2,000	4,000	2,300	17,000	4,375
Стандартное отклонение	136,189	1,344	2,949	0,920	14,351	3,749
Нижняя граница	0,000	1,219	1,598	1,549	5,632	1,861
Верхняя граница	254,658	3,907	7,496	3,389	34,334	9,359
Коэффициент асимметрии	1,687	0,531	2,09	0,633	1,986	1,428
Коэффициент эксцесса	2,535	1,035	8,054	0,141	5,649	2,335
Минимум	11	1	1	1	3,3	0,8
Максимум	628	5	18	5	85	18

Отметим, что нижняя и верхняя границы в *таблице 2* соответствуют так называемому интервалу «одна сигма» и определяются как разность и сумма среднего значения и среднеквадратического отклонения соответственно. В интервал, заключенный между этими границами, попадают наиболее вероятные значения анализируемой случайной величины, которые в случае нормального распределения составляют около 70% наблюдений выборки [7, 23].

Для обеспечения корректности анализа данные были распределены по классам однородности, число которых было определено по формуле Стерджеса (1) и оказалось равным семи. Гистограмма, представляющая собой графическую интерпретацию частотного распределения АГЗС по расстоянию до потребителей, представлена на *рисунке 1*.

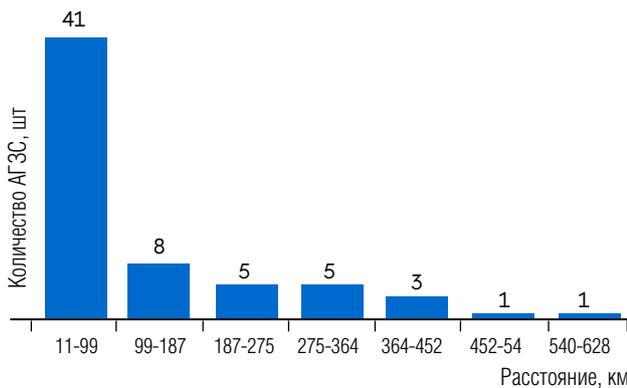


Рис. 1. Распределение потребителей по фактору «расстояние от компании до клиента».

На основе *рисунка 1* можно сделать предположение об экспоненциальности распределения расстояния до клиентов, что объясняется значительным увеличением их количества при приближении к городу. 41 АГЗС из 64 исследуемых (64%), находятся в пределах 100 км от компании. Среднее расстояние до клиентов составляет 118 км, половина всех значений укладывается в интервал от 11 км до 51,5 км, что говорит о сильном преобладании низких значений этого фактора, что подтверждается также и положительным значением коэффициента асимметрии. Наиболее вероятные значения этого показателя находятся в интервале (0–254), в него попадает 79% всех потребителей. Коэффициент эксцесса больше нуля, что говорит о достаточно хорошей прогнозируемости показателя относительно нормального распределения, поскольку большое число АГЗС сгруппировано в одном классе.

Аналогично была дана интерпретация частотных распределений числа АГЗС по остальным факторам.

Так, при анализе количества реализуемых видов топлива выяснилось, что наиболее часто встречаются АГЗС с двумя и четырьмя видами сжиженного газа (40%). Для показателя «производственные мощности» отмечено, что всего АГЗС имеют не более шести колонок. Анализируя данные о рейтинге клиентов, можно заметить не самую приятную статистику: чаще всего встречаются АГЗС с низким рейтингом (не выше 3 баллов), при этом среднее значение рейтинга находится в районе 2,5. Распределение значений объема цистерн похоже на экспоненциальное. Подавляющая часть клиентов (более 80%) имеет установленные цистерны суммарным объемом не выше 34 тыс. л. При анализе объемов суточных продаж выяснилось, что наибольшее число АГЗС характеризуется низкими продажами относительно остальных — менее 6 тыс. л газа в день. Всего шесть заправочных станций из 64 имеют среднесуточные продажи более 10 тыс. л (менее 10% от их общего числа).

После анализа исходных данных по выбранным показателям важно понять, насколько эти факторы влияют на итоговый результат («пригодность» клиента для сотрудничества). Для этой цели воспользуемся идеей построения перцептрона Розенблатта [1, 2].

На пятом этапе была построена нейросеть, состоящая из одного нейрона. В качестве функции активации нейрона (как и для всех других вариантов нейросетей, рассмотренных в работе), была взята логистическая функция в виде

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \quad (2)$$

Выбор логистической функции обусловлен ее непрерывностью, что обеспечивает гладкость в переходной области. В качестве функционала, определяющего корректность работы нейросети, использовалась величина ESS — остаточная сумма квадратов [7] между оценкой специалиста (Y) и оценкой, выдаваемой нейросетью (\hat{Y}):

$$ESS = \sum_{k=1}^N (Y_k - \hat{Y}_k)^2 \rightarrow \min. \quad (3)$$

В (3) суммирование проводится по всем записям базы данных, N — объем базы данных (число записей). Для входных факторов приняты обозначения x_1 – x_6 в порядке перечисления (*табл. 2*).

Вычисление оценок, определяемых нейросетью, проводилось на основе значения функции активации (2) от определяемых топологией нейросети линейных комбинаций значений входных переменных модели x_1-x_6 , а также любых переменных внутренних слоев нейросети. Так, например, для нейросети, состоящей из одного нейрона и имеющей шесть входных переменных, оценочное значение вероятности того, что клиент будет признан «пригодным» для сотрудничества, будет определяться как

$$\hat{Y} = \sigma(w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6) = \sigma(\Sigma), \quad (4)$$

где $w_i, i = 1, \dots, 6$ – весовые коэффициенты каждого из входных факторов, определяемые путем решения задачи (3);

Σ – значение линейной комбинации входных факторов, называемое сумматором.

Минимизация значения функционала (3) проводится путем изменения неизвестных весовых коэффициентов в сумматоре w_i . Поскольку в аналитической форме задача (3) решения не имеет, нахождение оценок весовых коэффициентов проводилось численным образом [24]. На рисунке 2 представлена схема реализованной модели из одного нейрона.

Модель простейшей сети включает значения входных факторов (круги слева на рис. 2), сумматор Σ из формулы (4), а также значение функции активации (2). Результат работы нейросети определяется долей корректных прогнозов – отношением числа верных срабатываний нейросети к общему объему базы данных N (в рассматриваемом случае $N = 64$). Степень затемнения стрелок отражает силу влияния соответствующего входного фактора на результат оценивания «пригодности» клиента, чем темнее соответствующая стрелка, тем сильнее влияние фактора.

Путем минимизации суммы квадратов отклонений вероятностей принятия решения о сотрудничестве с потребителем, оцененных экспертом компании, от прогнозируемых по нейросети ее значений получены оценки значений весовых ко-

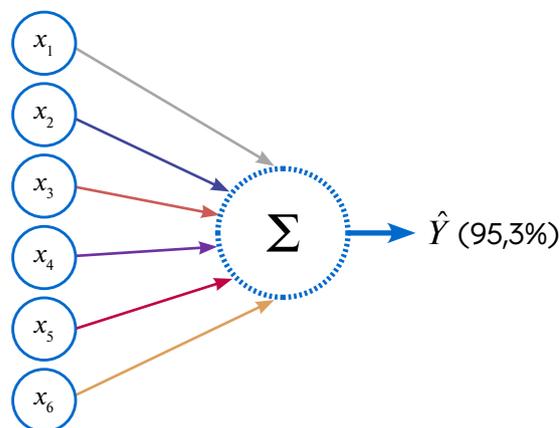


Рис. 2. Вид модели из одного нейрона.

эффициентов. В таблице 3 представлены значения оценок коэффициентов и результаты ранжирования анализируемых показателей по абсолютным величинам весовых коэффициентов.

Полученные значения весовых коэффициентов возможно интерпретировать: так, например, знак коэффициента говорит о характере зависимости. Из таблицы 3 видно, что только коэффициент w_1 оказался меньше нуля, что означает наличие отрицательной зависимости, т.е. с увеличением расстояния до потребителя вероятность принятия решения о сотрудничестве в среднем уменьшается. Остальные факторы имеют положительную зависимость с вероятностью принятия решения о сотрудничестве.

Значения коэффициентов по модулю означают «силу влияния» соответствующих факторов на принятие решения о сотрудничестве с тем или иным клиентом. Проанализируем полученные результаты.

1. Расстояние до АГЗС (фактор x_1) оказывает самое сильное влияние, его степень важности относительно всех остальных показателей составляет почти 60%. При этом оценка соответствующего ему весового коэффициента (w_1) упирается в границу допустимых значений (± 50), что означает

Таблица 3.

Оценки весовых коэффициентов нейрона и ранжирование показателей

Весовой коэффициент	w_1	w_2	w_3	w_4	w_5	w_6
Оценка коэффициента	-50,000	11,101	0,671	11,639	8,600	3,744
Степень важности, %	58,305	12,944	0,783	13,573	10,029	4,366
Ранг показателя	1	3	6	2	4	5

полное его доминирование при принятии решений. Достаточно логично, что компании не имеет смысла организовывать доставки на большие расстояния.

2. Рейтинг АГЗС (фактор x_4) занимает второе место, на него приходится чуть более 13% важности. Безусловно, чем выше репутация и оценки компании от клиентов, тем более надежной она является.
3. Разнообразие видов топлива на АГЗС (фактор x_2) стоит на третьем месте, немного уступая рейтингу (менее 13% важности). Важность данного фактора обуславливается более высокой финансовой стабильностью компании в случае наличия широкого ассортимента топлива для клиентов.
4. Объем цистерн (фактор x_5) располагается на четвертом месте. Влияние этого фактора подтверждается «удобством» выбранной АГЗС для сотрудничества. В какой-то степени этот показатель можно рассматривать как объем склада для торговой компании: он должен быть достаточным (как и запасы в нем), чтобы компания могла осуществлять свою деятельность без почасовой доставки продукции. Для АГЗС ситуация аналогичная. Путем оптимизации можно минимизировать необходимый объем цистерн, но доставка продукции при помощи систем типа just-in-time [25] может быть достаточно сложной по причине специфики перевозимых грузов, поэтому будет важно иметь запасные емкости для хранения газа.
5. Суточный объем продаж газа (фактор x_6) находится на предпоследнем месте (менее 5% важности), что несколько удивляет. Влияние этого показателя должно быть очевидным, ведь чем больше оборот компании, тем выше ее стабильность. Однако важно понимать, что все компании (равно как и АГЗС) находятся в разных ситуациях. Так, для мощных станций на шесть и более колонок, располагающихся в центре города, высокий, с первого взгляда, оборот может на самом деле оказаться достаточно низким по сравнению с близко расположенными к нему другими клиентами.
6. Производственная мощность АГЗС (фактор x_3) занимает последнее место (важность менее 1%). Как и предыдущий показатель, он не имеет четких рамок, поэтому как фактор он не является сильным. В целом большее число колонок является возможностью для развития компании, что может сыграть свою роль в долгосрочной перспективе.

Отметим, что полученные при помощи нейросети оценки корректно отражают мыслительный

процесс специалиста, который занимается этим вручную. Анализируя результаты работы нейросети, следует отметить, что из 64 записей базы данных о клиентах компании нейрон ошибся только в трех, что составляет 4,7% ошибок. Для понимания причин этих ошибок нужно рассмотреть результаты и выявить особенности этих АГЗС. Во всех трех случаях оценка вероятности «надежности» клиента, выдаваемая нейросетью, была более 0,9, т.е. она была более чем уверена в их «пригодности» для сотрудничества.

Первая АГЗС обладает невысокими значениями почти всех показателей, однако она имеет высокий рейтинг, что, скорее всего и послужило причиной такой оценки. Очевидно, что рейтинги, предоставляемые специальными сервисами – самый ненадежный показатель, поскольку в целом ряде случаев они или нерепрезентативны из-за малого числа усредняемых оценок, или некорректны из-за использования определенных схем «накручивания» нужных значений.

Вторая АГЗС обладает хорошими значениями показателей, однако специалист компании отметил, что с данной станцией они не сотрудничают, поскольку другой филиал находится ближе к клиенту, и взаимодействие осуществляется через них.

Третья АГЗС была оценена положительно, но специалист негативно оценил опыт работы с этой компанией потому, что станция находится в процессе запуска, данные по ней противоречивы, и нужно больше времени для подписания договоров. Такая ситуация не является ошибкой, однако в свете повышения распознавательных способностей нейросети можно порекомендовать добавить проверку на время работы АГЗС на рынке.

На шестом этапе для исключения ошибочных срабатываний нейросети было решено рассмотреть более сложные варианты, добавив внутренние слои из разного числа нейронов, наиболее подходящим вариантом оказалась топология с семью нейронами 4-2-1 (рис. 3).

Выбранная топология позволила успешно описать процесс взаимодействия выбранных показателей, при этом расхождений между оценками специалиста и результатами работы нейросети выявлено не было. Более темные стрелки связей нейронов в сети соответствуют третьему нейрону первого слоя и шестому нейрону второго, при этом разница влияния значений нейронов второго слоя на выход нейросети составляет лишь 18,4%. Также

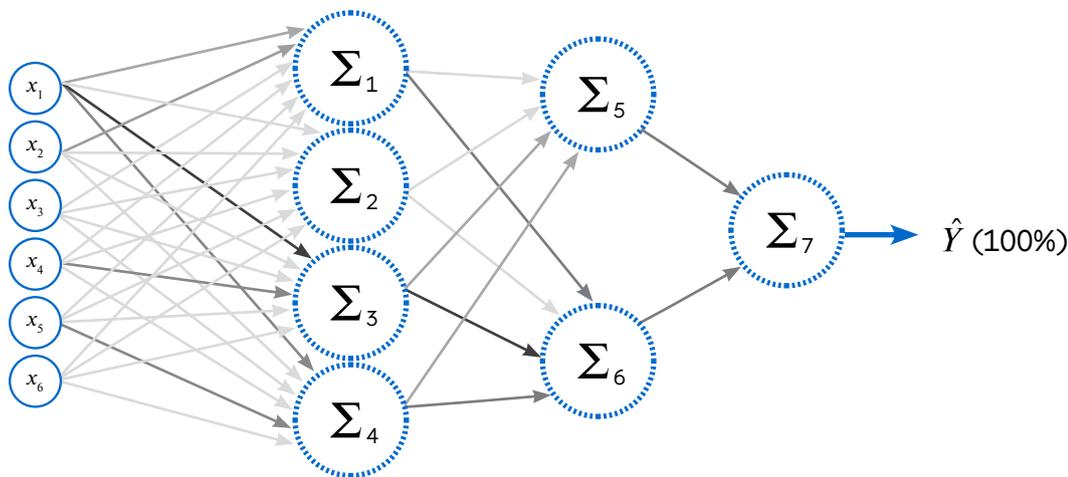


Рис. 3. Вид модели нейросети по топологии 4-2-1.

отметим, что значения пятого нейрона оказывают отрицательное влияние на результат, а шестого – положительное. К сожалению, этот факт нельзя использовать для интерпретации результатов (как это было для нейросети из одного нейрона), поскольку значения каждого из нейронов внутренних слоев складываются под влиянием предыдущих слоев нелинейным образом из-за выбранной функции активации в виде логистической функции (2).

В целях обеспечения адекватности получаемых результатов, помимо модели нейросети были построены классификации клиентов при помощи логистической регрессии, деревьев принятия решений и случайного леса. Для разработки и идентификации соответствующих моделей был использован свободно распространяемый программный комплекс Orange Data Mining интеллектуального анализа данных [19].

Для построения наилучшей модели логистической регрессии был выбран метод максимального правдоподобия как наиболее часто применяемый в подобных ситуациях. В рамках данного исследования был рассмотрен стандартный вариант линейной по параметрам и входным факторам регрессионной модели.

Что касается построения модели при помощи методов дерева решений и случайного леса, то были рассмотрены различные варианты деревьев, качество классификаций, строимых на их основе, определялось при помощи стандартного показателя – меры F_1 [4, 19, 22]. При этом для обеспечения построения эффективных деревьев классификации были рассмотрены деревья с различными параметрами, результаты представлены в таблице 4 и таблице 5. Кроме того, в таблице 6 приведены результаты сопоставления классификаций по наилучшим результатам полученных моделей.

В таблице 4 представлены результаты оценивания качества моделей, построенных по методу дерева решений. Здесь использованы следующие обозначения: a – минимальное число элементов в одном листе дерева; b – количество элементов в листе, при котором не производится дальнейшее разбиение; F_1 – значение меры качества классификаций. В таблице 4 и таблице 5 жирным шрифтом выделены варианты классификаций, которые оказываются наилучшими с точки зрения максимального значения меры качества. В таблице 5 представлены данные результатов оценивания качества моделей, построенных по методу случайного леса. В скобках после меры F_1 указано число деревьев соответствующего случайного леса.

Таблица 4.

Параметры модели дерева решений и качество получаемых классификаций

a	2		4		6	8	10	
b	2, 4, 6	8	10	4, 6, 8	10	6, 8, 10	8, 10	10
F_1	0,876	0,891	0,857	0,889	0,857	0,842	0,842	0,844

Таблица 5.

Параметры моделей случайного леса и качество получаемых классификаций

<i>a</i>	2	2	2	2	2	4	4	4	4	6	6	6	8	8	10
<i>b</i>	2	4	6	8	10	4	6	8	10	6	8	10	8	10	10
$F_1(10)$	0,840	0,840	0,855	0,855	0,870	0,874	0,889	0,859	0,874	0,889	0,859	0,874	0,859	0,874	0,874
$F_1(50)$	0,904	0,904	0,904	0,889	0,889	0,874	0,904	0,874	0,874	0,889	0,874	0,874	0,874	0,874	0,874
$F_1(100)$	0,887	0,887	0,887	0,887	0,887	0,874	0,889	0,889	0,874	0,874	0,889	0,874	0,889	0,874	0,871

Отметим, что при некоторых комбинациях параметров случайного леса результирующие классификации оказываются одинаковыми, что отражается в значениях F_1 -меры. Кроме того, дальнейшее увеличение числа деревьев в лесе не приводит к изменениям в значениях меры качества классификаций, что является логичным следствием усреднения свойств получаемых деревьев.

Результаты классификации по перечисленным методам представлены в *таблице 4*, где обозначение «1» использовалось для «подходящих» клиентов, с которыми компании выгодно сотрудничать, а «0» – для «неподходящих». При этом «наблюдаемые значения» соответствуют оценкам специалиста, а «прогнозируемые» – результатам классификаций по рассматриваемым методам.

Из *таблицы 6* видно, что применение логистической регрессии привело к шести ошибкам (9,4% ошибок), три из которых возникли для «подходящих» клиентов, а другие три – для «неподходящих»; при использовании метода дерева решений число ошибок для «подходящих» клиентов составило два, а «неподходящих» – пять (итого – 10,9% ошибок); метод случайного леса привел к появлению

четырёх ошибок для «неподходящих» клиентов (всего 7,8% ошибок). В целом можно отметить, что методы, альтернативные нейросетевому, не обеспечили полного совпадения оценок эксперта и результатов классификации. Это позволяет сделать вывод о том, что для решения рассматриваемой задачи модель нейросети является наиболее подходящей для оценивания качества клиентов.

Необходимо понимать, что для проверки корректности работы нейронной сети недостаточно использовать только набор данных, который имеется на текущий момент. Важно обеспечить сохранение ее прогностических свойств для новых порций данных, что со статистической точки зрения означает обеспечить адекватность модели [6]. Поэтому после окончания обучения нейросети для проверки ее работоспособности были взяты десять новых АГЗС, контракты с которыми были заключены в процессе проведения исследования, и которые, естественно, не попали в первоначальную базу данных. Результаты работы нейросети (в виде оценок вероятности «пригодности» клиента и окончательной оценки) были сопоставлены с оценками специалиста (*табл. 7*).

Таблица 6.

Оценка качества методов классификации

		Прогнозируемые по логистической регрессии		Прогнозируемые по дереву решений		Прогнозируемые по методу случайного леса		Σ
		0	1	0	1	0	1	
Наблюдаемые значения	0	20	3	18	5	19	4	23
	1	3	38	2	39	0	41	41
	Σ	23	41	20	44	19	45	64

Таблица 7.

Сравнение оценок нейросети и эксперта по новым клиентам

№ п/п	1	2	3	4	5	6	7	8	9	10
Оценка эксперта	1	0	1	1	1	1	1	1	1	1
Оценка вероятности	0,954	0,121	0,851	0,884	0,732	0,999	0,411	0,970	0,804	0,925
Оценка нейросети	1	0	1	1	1	1	0	1	1	1

Нейросеть верно оценила девять позиций из десяти, ошибка совершена с седьмой станцией, погрешность в оценке вероятности составляет менее одной десятой (0,411 против 0,500, что было бы достаточно для признания АГЗС пригодной для сотрудничества). Эта станция обладает высоким рейтингом (4), реализует пять видов топлива и имеет среднюю мощность (четыре колонки). Однако она расположена в 149 км от компании и имеет установленные цистерны относительно невысокой емкости (порядка 16 тыс. л). Объем продаж газа составляет менее 6 тыс. л, т.е. данная АГЗС весьма средняя, но неоднозначная по характеристикам.

Нейросеть оценила, что сотрудничество с таким клиентом не будет выгодным, но специалист компании решил, что она подходит. Это противоречие может быть признаком того, что для данной АГЗС, возможно, следует провести более тщательное изучение ее возможностей для обеспечения успешного сотрудничества с компанией. При накоплении определенного числа таких «спорных» ситуаций может быть принято решение о переобучении нейросети на более обширной базе данных, либо, если результаты продолжают не соответствовать действительности, то может потребоваться более радикальное действие – изменение топологии нейросети и полное ее переобучение.

На седьмом этапе для самостоятельного анализа клиента на соответствие требованиям компании можно использовать список пороговых значений для каждого показателя, который поможет быстро оценить, насколько клиенты со значениями факторов, близкими к средним, подходят для дальнейшего анализа.

Для построения такого перечня при помощи разработанной нейросети были поочередно определены диапазоны значений показателей, при которых «пригодность» клиента сохраняется, при этом значения всех остальных характеристик были зафиксированы на средних уровнях (табл. 2). На ос-

новании весовых коэффициентов нейросеть самостоятельно рассчитывает значения, при которых оценка вероятности \hat{Y} (оценка «надежности клиента») будет в промежутке от 0,5 до 1, что соответствует принятию решения о том, что клиент подходит для сотрудничества. Таким образом, составим таблицу пороговых значений (табл. 8), которая может использоваться в качестве подсказки для специалиста при проведении оценивания степени пригодности клиента.

Сопоставляя данные в таблице 2 и таблице 8, можно заключить, что результаты моделирования, полученные в ходе исследования, позволяют говорить о корректности применения нейросетей для задач выбора покупателей в оптово-розничной торговле продуктами нефтегазовой сферы.

4. Обсуждение результатов

В процессе апробации разработанной методики возникла необходимость в построении дополнительной интерпретации полученных результатов, не только в части получения аналитических и статистических выводов, но и для разработки наглядных и конкретных рекомендаций по взаимо-

Таблица 8.

Перечень пороговых значений показателей клиентов

Наименование фактора	Допустимое значение	
	min	max
Расстояние до клиента, км	–	156
Широта ассортимента, ед.	2	–
Производственная мощность, ед.	1	–
Рейтинг АГЗС (по 5-балльной шкале)	1,1	–
Емкость цистерн, тыс. л	20	–
Объем продаж, тыс. л	5,6	–

действию с клиентами компании. Так, например, интерес представляет построение географической интерпретации полученных результатов, например, изобразив на карте примерную зону покрытия компании и проанализировать ее. Радиус этой зоны, согласно *таблице 8*, составляет 156 км.

Предельное значение расстояния можно назвать важнейшим показателем, поскольку он основан на факторе, имеющем наибольший вес в нейросети. Оно определяет, насколько далеко расположены потенциальные клиенты и, отчасти, насколько удобен доступ к автозаправочным газовым станциям (АГЗС) для них. Эта величина играет ключевую роль для компании, поскольку она напрямую влияет на стоимость транспортировки газа до потребителей, что, в свою очередь, может существенно повлиять на их собственные расходы.

Для уточнения показателя «расстояние до клиента» можно добавить новую характеристику, учитывающую направление расположения клиента относительно местоположения компании. Однако при этом важно понимать, что весовой коэффициент нового показателя будет «отнимать» часть силы старого, так как они учитывают похожие особенности клиентов.

Другой вариант решения данной проблемы состоит в замене имеющегося показателя на четыре новых, каждый из которых является расстоянием в определенном направлении (например, север-юг-запад-восток). Тогда расстояние до каждой АГЗС будет учитываться по одному—двум отдельным показателям (например, городские заправочные станции будут преимущественно расположены по направлению на запад, север или юг вследствие специфики географии г. Новосибирска). Такой подход позволит определить более точные пограничные значения расстояний и поможет в более четком оценивании зоны покрытия, которая, предположительно, будет вытянутой в разные стороны. Например, максимально возможное расстояние на восток, в сторону Новосибирска, будет заметно больше, чем на запад, так как большая часть города, и, соответственно, возможные клиенты компании, расположены на восточном берегу реки Обь.

Важно понимать, что определение переходного значения расстояния является сложной задачей и может зависеть от множества факторов (транспортная инфраструктура, плотность населения, топография местности и др.). Поэтому при принятии решения о сотрудничестве с АГЗС необходи-

мо учитывать эти факторы и проводить подробный анализ рынка и соответствующей инфраструктуры.

Пороговое количество видов топлива важно для заправочных станций, поскольку оно влияет на потребительский сервис и эффективность управления запасами топлива. Этот показатель может зависеть от разных факторов, таких как местоположение АГЗС и потребности местных жителей. Обычно на удаленных станциях предлагается не более двух—трех видов топлива для автомобилей, что соответствует найденному пороговому значению в два вида топлива. Если станция осуществляет заправку газом, то одновременно осуществляется и его реализация в баллонах, что может быть удобным для владельцев автомобилей, которые используют газ как в качестве топлива, так и для бытовых нужд.

Нижнее пороговое значение мощности АГЗС оказалось равным единице, что, скорее всего, вызвано малой степенью его важности (0,783%, *табл. 3*). Однако если на станции слишком мало колонок, то она может оказаться недостаточной для обеспечения потребностей всех пользователей. С другой стороны, слишком высокая мощность может привести к избыточным затратам на строительство и обслуживание АГЗС. Поэтому перед выбором мощности АГЗС рекомендуется провести анализ потребностей пользователей, доступности газопроводов и других факторов, чтобы выбрать оптимальную мощность, которая будет соответствовать требованиям всех заинтересованных сторон и обеспечивать максимальную эффективность и экономичность использования АГЗС.

Минимально возможный рейтинг АГЗС составил 1,1 балла по пятибалльной шкале, что оказывается весьма низким и является следствием низкого уровня удовлетворенности клиентов, выраженных в отзывах об АГЗС. Для получения более объективной картины о качестве сервиса и уровне удовлетворенности клиентов на АГЗС, необходимо изучить отзывы из 2ГИС и других источников, используя при этом средневзвешенную оценку, учитывающую репутацию и надежность источников, а также их количество.

Пороговое значение объема установленных на станциях цистерн равно 20 тыс. л, что вполне соответствует объемам средних АГЗС (*табл. 2*, среднее равно 19,983 тыс. л). Однако, если расходы заметно превышают ожидания, то установленных емкостей может хватать не более чем на день, что повлечет за собой частые дозаправки по несколько раз в день,

что может потребовать установки дополнительных цистерн для увеличения общей емкости.

Минимальные допустимые дневные продажи на АГЗС – это важный показатель, который определяет минимальный объем топлива, который должен быть продан (и, соответственно, быть в наличии на момент продажи), чтобы станция могла оставаться рентабельной. Для среднестатистических АГЗС минимальные допустимые дневные продажи составляют 5600 литров. Однако такая величина является достаточно низкой для городских станций, которые обслуживают большое количество автомобилей. В то же время для АГЗС на окраинах города или в малонаселенных районах, где объем продаж топлива ниже, такое значение будет более чем достаточным.

В дальнейших модификациях нейросети имеет смысл заменить этот показатель на более объективный, такой как отношение величины продаж к численности населения того района, в котором располагается АГЗС, чтобы учитывать потенциальный спрос на топливо в конкретном районе. Кроме того, можно рассчитать отдельные показатели для города и области, учитывая особенности регионального рынка топлива.

Дополнительный интерес представляет также исследование при помощи нейросети того, как могут меняться величины пороговых значений при изменении не только одного, но и двух (трех и т.д.) других показателей одновременно. Пример подобного расчета, проведенного при помощи методов регрессионного анализа для построения соответствующих зависимостей [6], представлен на *рисунке 4*.

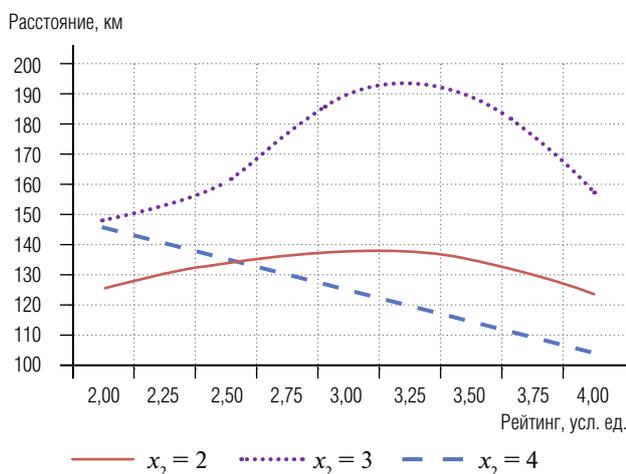


Рис. 4. Сравнение верхних пороговых значений расстояния в зависимости от рейтинга и числа видов топлива.

Анализируя *рисунок 4*, можно говорить о том, что характер изменения максимально возможного расстояния существенно меняется: для потребителей, реализующих четыре вида топлива ($x_2 = 4$) с увеличением рейтинга расстояние уменьшается, тогда как при трех или четырех видах топлива – сначала увеличивается, а потом уменьшается. Более того, можно определить максимально допустимое расстояние, которое для клиентов с $x_2 = 2$ равно 193 км, а для клиентов с $x_2 = 3$ – 138 км.

Полученные пороговые значения могут использоваться для быстрой оценки клиентов при заключении договора. При дальнейшей разработке нейросети факторы могут изменяться как количественно, так и качественно, поэтому своевременное обновление пула входных факторов и содержимого базы данных будет позволять получать более корректные оценки клиентов.

Заключение

Результатом применения авторской методики является построение модели принятия решения о выборе покупателей, с которыми компания планирует дальнейшее взаимовыгодное сотрудничество. С этой целью был сформирован пул факторов, характеризующих потребителей и их покупательское поведение. На основе базы данных о покупателях компании и экспертных оценок, определяющих их надежность, были разработаны нейронные сети, позволяющие оценивать перспективность сотрудничества с клиентами. При их помощи была решена задача ранжирования факторов, характеризующих потребителей, относительно степени и характера их влияния на процесс принятия решения о надежности того или иного клиента.

По результатам обучения нейросетей была выбрана сеть «наилучшей» топологии, которая обеспечила корректное прогнозирование для всех записей базы данных. Сравнение результатов работы этой нейросети с классификациями, построенными на основе других методов интеллектуального анализа данных, позволило сделать вывод о наилучшей адекватности модели нейросети для решения рассматриваемой задачи.

Построенная нейросеть была использована для определения пороговых значений входных факторов модели. Это позволило разработать рекомендации для сотрудников компании по отбору предложений о сотрудничестве с потребителями.

Дальнейшее совершенствование предложенной методики может заключаться в расширении пула входных факторов за счет привлечения новых характеристик потребителей, в том числе и предлагаемых в настоящей работе, а также путем расщепления имеющихся факторов на несколько составляющих, каждая из которых характеризует определенную специфику потребителей. Кроме того, для нейросетей с большим числом входных переменных име-

ет смысл рассматривать более сложные топологии, которые могут включать не только дополнительные внутренние слои, но и обратные связи.

Разработанная методика вследствие своей универсальности может быть рекомендована для решения различных классификационных задач не только в сфере логистики, но и для широкого спектра проблем экономического и управленческого характера. ■

Литература

1. Розенблатт Ф. Принципы нейродинамики: Перцептроны и теория механизмов мозга. Пер. с англ. М: Мир, 1965.
2. Rumelhart D.E., Hinton G.E., Williams R.J. Learning internal representations by error propagation // *Parallel Distributed Processing*, Cambridge, MA, MIT Press, 1986. Vol. 1. P. 318–362. <https://doi.org/10.1016/B978-1-4832-1446-7.50035-2>
3. Галушкин А.И. Нейронные сети: основы теории. М.: Горячая линия – Телеком, 2010.
4. Hinton G., Le Cun Y., Bengio Y. Deep learning // *Nature*. 2015. Vol. 521. P. 436–444. <https://doi.org/10.1038/nature14539>
5. Севастьянов Л.А., Щетинин Е.Ю. О методах повышения точности многоклассовой классификации на несбалансированных данных // *Информатика и ее применения*. 2020. Т. 14. №. 1. С. 63–70. <https://doi.org/10.14357/19922264200109>
6. Тимофеев В.С., Фаддеенков А.В., Шеколдин В.Ю. Эконометрика. М.: Юрайт, 2016.
7. Цой М.Е., Шеколдин В.Ю. Маркетинговые исследования: методы анализа маркетинговой информации. Новосибирск: Изд-во НГТУ, 2021.
8. Breiman L. Random forests // *Machine Learning*. 2001. Vol. 45. No. 1. P. 5–32. <https://doi.org/10.1023/A:1010933404324>
9. Чистяков С.П. Случайные леса: обзор // *Труды Карельского научного центра РАН*. 2013. № 1. С. 117–136.
10. Abbate R., Manco P., Caterino M., Fera M., Macchiaroli M. Demand forecasting for delivery platforms by using neural network // *IFAC-Papers OnLine*. 2022. Vol. 55. No. 10. P. 607–612. <https://doi.org/10.1016/j.ifacol.2022.09.465>
11. Данильченко М.Н., Муравник А.Б. Нейросетевой подход к построению маршрута в автоматизированной системе управления специального назначения // *Наукоемкие технологии в космических исследованиях Земли*. 2021. Т. 13. № 1. С. 58–66.
12. Sustrova T. A suitable artificial intelligence model for inventory level optimization // *Trends Economics and Management*. 2016. Vol. 10(25). P. 48–55. <https://doi.org/10.13164/trends.2016.25.48>
13. Михайлин Д.А. Нейросетевой алгоритм безопасного облета воздушных препятствий и запрещенных наземных зон // *Научный вестник МГТУ ГА*. 2017. Т. 20. № 4. С. 18–24. <https://doi.org/10.26467/2079-0619-2017-20-4-18-24>
14. Hughes A. *Boosting Response with RFM*. New York: Marketing Tools, 1996.
15. Griffin J. *Customer loyalty: how to earn it, how to keep it*. San Francisco, CA: Jossey Bass, 1996.
16. Guo Li. A research on influencing factors of consumer purchasing behaviors in cyberspace // *International Journal of Marketing Studies*. 2011. Vol. 3. No. 3. P. 182–188. <https://doi.org/10.5539/ijms.v3n3p182>
17. Цой М.Е., Шеколдин В.Ю., Лежнина М.Н. Построение сегментации на основе модифицированного RFM-анализа для повышения лояльности потребителей // *Российское предпринимательство*. 2017. Т. 18. № 21. С. 3113–3134. <https://doi.org/10.18334/rp.18.21.38506>
18. Сондерс М., Льюис Ф., Торнхилл Э. Методы проведения экономических исследований. Пер. с англ. М.: ЭКСМО, 2006.
19. Orange: Data mining toolbox in Python / J. Demsar [et al.] // *Journal of Machine Learning Research*. 2013. Vol. 14. P. 2349–2353.
20. Sturges H. The choice of a class-interval // *Journal of the American Statistical Association*. 1926. Vol. 21. P. 65–66. <https://doi.org/10.1080/01621459.1926.10502161>
21. Neural networks: An overview of early research, current frameworks and new challenges / A. Prieto [et al.] // *Neurocomputing*. 2016. Vol. 214. P. 242–268. <https://doi.org/10.1016/j.neucom.2016.06.014>
22. Аптон Г. Анализ таблиц сопряженности. М.: Финансы и статистика, 1982.
23. Кокрен У. Методы выборочного исследования. М.: Статистика, 1976.
24. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы. Москва: Лаборатория знаний, 2020.
25. Лайкер Дж. Дао Тойота: 14 принципов менеджмента ведущей компании мира. М.: Альпина Бизнес Букс, 2005.

Об авторах

Назаркина Валерия Александровна

к.э.н., доц.;

доцент, кафедра маркетинга и сервиса, Новосибирский государственный технический университет, Россия, 630073, г. Новосибирск, пр-т К. Маркса, д. 20;

E-mail: valeria71@bk.ru

ORCID: 0000-0002-2207-5228

Щеколдин Владислав Юрьевич

к.т.н.;

доцент, кафедра маркетинга и сервиса, Новосибирский государственный технический университет, Россия, 630073, г. Новосибирск, пр-т К. Маркса, д. 20;

E-mail: schekoldin@corp.nstu.ru

ORCID: 0000-0001-8016-1282

Neural network technologies in supply chain management: Consumer selection technique

Valeriya A. Nazarkina

E-mail: valeria71@bk.ru

Vladislav Yu. Shchekoldin

E-mail: schekoldin@corp.nstu.ru

Novosibirsk State Technical University (NSTU), Novosibirsk, Russia

Abstract

The supply chain management's effectiveness depends, among other things, on the selection and coordinated interaction with product consumers. This article is devoted to the development of a method for selecting a consumer in the regional wholesale and retail fuel market. The methodological basis of the study is the theory of statistical analysis and neural networks. The main tool for developing the methodology was neural network technologies, with the help of which it is most likely possible to correctly estimate the boundaries for indicators' values that characterize consumers and reflect their history of purchasing behavior, to select potential clients and the possibility of further cooperation with existing ones. The information base for the work is the data on consumers of a given company's products, data from the 2GIS electronic directory, as well as the results of the primary statistical analysis and forecasts made based on neural networks of various topologies. The author presents his methodology for selecting a consumer. It has the potential for development and implementation for solving a number of other management problems. As part of the testing, the best configuration (topology) of the neural network was determined, and standard

values of entry barriers when consumer choice accomplished were assessed. The methodology we developed was tested using the example of a company operating in the wholesale and retail fuel market in Novosibirsk and the Novosibirsk region. When verifying the neural network model, the quality of client classification was compared based on logistic regression, decision tree and random forest models and we found that the neural network approach provides the best results for assessing the degree of client suitability. As a result of testing the methodology, recommendations for improving neural network models were developed, including expanding the set of factors that determine the characteristics of consumers, as well as optimizing the internal structure of neural networks.

Keywords: neural networks, supply chain management, distribution logistics, consumer selection, neural network training

Citation: Nazarkina V.A., Shchekoldin V.Yu. (2024) Neural network technologies in supply chain management: Consumer selection technique. *Business Informatics*, vol. 18, no. 4, pp. 7–24. DOI: 10.17323/2587-814X.2024.4.7.24

References

1. Rosenblatt F. (1965) *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Moscow: Mir (in Russian).
2. Rumelhart D.E., Hinton G.E., Williams R.J. (1986) Learning internal representations by error propagation. *Parallel Distributed Processing, Cambridge, MA, MIT Press*, vol. 1, pp. 318–362. <https://doi.org/10.1016/B978-1-4832-1446-7.50035-2>
3. Galushkin A.I. (2010) *Neural networks: basic theory*. Moscow: Hotline –Telecom (in Russian).
4. Hinton G., Le Cun Y., Bengio Y. (2015) Deep learning. *Nature*, vol. 521, pp. 436–444. <https://doi.org/10.1038/nature14539>
5. Sevastyanov L.A., Shchetinin E.Yu. (2020) On methods for increasing the accuracy of multi-class classification on unbalanced data. *Informatics and its applications*, vol. 14, no. 1, pp. 63–70 (in Russian). <https://doi.org/10.14357/19922264200109>
6. Timofeev V.S., Faddeenkov A.V., Shchekoldin V.Yu. (2016) *Econometrics*. Moscow: YURAYT (in Russian).
7. Tsoi M.E., Shchekoldin V.Yu. (2021) *Marketing research: methods for analyzing marketing information*. Novosibirsk: Publishing house of NSTU (in Russian).
8. Breiman L. (2001) Random Forests. *Machine Learning*, vol. 45, no. 1, pp. 5–32. <https://doi.org/10.1023/A:1010933404324>
9. Chistyakov S.P. (2013) Random forests: a review. *Proceedings of the Karelian Scientific Center of the Russian Academy of Sciences*, no. 1, pp. 117–136 (in Russian).
10. Abbate R., Manco P., Caterino M., et al. (2022) Demand forecasting for delivery platforms by using neural network. *IFAC-Papers OnLine*, vol. 55, no. 10, pp. 607–612. <https://doi.org/10.1016/j.ifacol.2022.09.465>
11. Danilchenko M.N., Muravnik A.B. (2021) Neural network approach to route construction in a special-purpose automated control system. *High-tech technologies in space exploration of the Earth*, vol. 13, no. 1, pp. 58–66 (in Russian). <https://doi.org/10.36724/2409-5419-2021-13-1-58-66>
12. Sustrova T. (2016) A suitable artificial intelligence model for inventory level optimization. *Trends Economics and Management*, vol. 10(25), pp. 48–55. <https://doi.org/10.13164/trends.2016.25.48>
13. Mikhailin D.A. (2017) Neural network algorithm for safe flight around air obstacles and prohibited ground zones. *Scientific Bulletin of MSTU GA*, vol. 20, no. 4, pp. 18–24 (in Russian). <https://doi.org/10.26467/2079-0619-2017-20-4-18-24>
14. Hughes A. (1996) *Boosting Response with RFM*. New York: Marketing Tools.
15. Griffin J. (2002) *Customer loyalty: how to earn it, how to keep it*. San Francisco, CA: Jossey-Bass.
16. Guo Li. (2011) A research on influencing factors of consumer purchasing behaviors in cyberspace. *International Journal of Marketing Studies*, vol. 3, no. 3, pp. 182–188. <https://doi.org/10.5539/ijms.v3n3p182>
17. Tsoi M.E., Shchekoldin V.Yu., Lezhnina M.N. (2017) Construction of segmentation based on modified RFM analysis to increase consumer loyalty. *Russian Entrepreneurship*, vol. 18, no. 21, pp. 3113–3134 (in Russian). <https://doi.org/10.18334/rp.18.21.38506>
18. Saunders M., Lewis F., Thornhill E. (2006) *Methods of conducting economic research*. Moscow: EKSMO (in Russian).
19. Demsar J., Curk T., Erjavec A., et al. (2013) Orange: Data mining toolbox in Python. *Journal of Machine Learning Research*, vol. 14, pp. 2349–2353.

20. Sturges H. (1926) The choice of a class-interval. *Journal of the American Statistical Association*, vol. 21, pp. 65–66. <https://doi.org/10.1080/01621459.1926.10502161>
21. Prieto A., Prieto B., Ortigosa E.M., et al. (2016) Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, vol. 214, pp. 242–268. <https://doi.org/10.1016/j.neucom.2016.06.014>
22. Upton G. (1982) *Analysis of contingency tables*. Moscow: Finance and Statistics (in Russian).
23. Cochran W. (1976) *Sampling methods*. Moscow: Statistics (in Russian).
24. Bakhvalov N.S., Zhidkov N.P., Kobelkov G.M. (2020) *Numerical methods*. Moscow: Knowledge Laboratory (in Russian).
25. Liker J. (2005) *Dao Toyota: 14 principles of management of the world's leading company*. Moscow: Alpina Business Books (in Russian).

About the authors

Valeriya A. Nazarkina

Cand. Sci. (Econ.), Assoc. Prof.;

Associate Professor, Marketing and Service Department, Novosibirsk State Technical University (NSTU), 20, Karl Marx Ave., Novosibirsk 630092, Russia;

E-mail: valeria71@bk.ru

ORCID: 0000-0002-2207-5228

Vladislav Yu. Shchekoldin

Cand. Sci. (Tech.);

Associate Professor, Marketing and Service Department, Novosibirsk State Technical University (NSTU), 20, Karl Marx Ave., Novosibirsk 630092, Russia;

E-mail: schekoldin@corp.nstu.ru

ORCID: 0000-0001-8016-1282