

На правах рукописи

ПИВКИН КИРИЛЛ СЕРГЕЕВИЧ

**МОДЕЛИРОВАНИЕ ПОКУПАТЕЛЬСКОГО СПРОСА НА
ПРЕДПРИЯТИЯХ РОЗНИЧНОЙ ТОРГОВЛИ НА ОСНОВЕ МЕТОДОВ
МАШИННОГО ОБУЧЕНИЯ**

08.00.13 – Математические и инструментальные методы экономики

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата экономических наук

Ижевск 2018

I. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. Конъюнктура розничной торговли является ключевым индикатором развития экономики страны. Как одна из самых подвижных сфер экономики розничная сфера ярко отражает работу экономических законов как микро-, так и макроуровня. После испытаний мирового финансового кризиса и современного структурно-экономического кризиса в России конкурентоспособность отечественных предприятий розничной торговли стала, прежде всего, определяться способностью оптимизировать внутренние бизнес-процессы и потоки товародвижения. Это приводит к созданию новых методов и подходов в политике организации розничного бизнеса, в частности, в области планирования, прогнозирования и организации товародвижения.

Стохастичность спроса накладывает большие ограничения на работу розничной компании. При отсутствии системы прогнозирования должного качества страдает большинство областей управления розничного предприятия, например, область управления запасами, где низкая точность решения может привести к снижению конкурентоспособности. При этом, для торговых компаний с узким ассортиментом наиболее популярных брендов подобная проблема менее актуальна – она решается с применением дешевого труда менеджеров по ежедневному контролю низко вариативного спроса. С увеличением количества товарных позиций, увеличением торговых площадей, точек продаж и масштабов торговли обеспечить качественное прогнозирование спроса путем прямых калькуляций невозможно на практике. Поэтому важным является совершенствование текущих технологий прогнозирования и планирования покупательского спроса, автоматизации этих бизнес-процессов, причем на основе продвинутого статистического моделирования.

Современные методы прикладной статистики получили обширное применение во многих отраслях экономики ввиду технологического бума. Развитие скорости и эффективности вычислительных алгоритмов на сегодняшний день позволяет обрабатывать большие массивы данных (Big Data) даже на персональных компьютерах с помощью открытого программного обеспечения. Имея основания для разработки эффективной платформы моделирования покупательского спроса, ставится задача о повышении точности прогнозирования. Исходя из наличия проработанных в теории и на практике методов – специальных методов линейной регрессии, деревьев решений и их композиций, метода опорных векторов, нейронных сетей, градиентного бустинга и т.п. – задача высокоточного моделирования потребности представляется достижимой. Наиболее качественный прогноз позволяет решить проблему стохастичности процесса товародвижения. Затем результат моделирования предприятие может использовать в процессах управления запасами, в планировании маркетинговых мероприятий, в динамичном управлении ценами и других областях управления.

Для построения развитой системы прогнозирования покупательского спроса ощущается острая потребность в реализации модели прогнозирования на

основании методов современной теории эконометрики, статистического и машинного обучения, а также элементов теории экономики торговли и управления запасами, при этом имея конкретную форму в виде реализованного программно-вычислительного комплекса, интегрированного в информационные системы предприятия. Наличием данной потребности и обусловлена актуальность настоящего диссертационного исследования.

Степень разработанности проблемы определена как большим количеством литературных источников и публикаций по экономике торговых организаций, эконометрике и теории машинного обучения, так и достаточным количеством проработанных концепций для математического прогнозирования и моделирования, большая часть которых не универсальна и сложна во внедрении для решения бизнес-задач.

Основные понятия экономики торговых организаций даны в трудах отечественных авторов и авторов ближнего зарубежья: Р. П. Валевича, Г. А. Давыдовой, М. С. Абрютиной, Л. А. Брагина, Т. П. Данько, Л. А. Козерод, Т. М. Безбородова, М. Б. Дюжева, В. В. Лукинского. В данных источниках раскрыты принципы работы торговых организаций, в частности компаний розничной торговли, а также описана ключевая роль управления товарными запасами в системе менеджмента торговой организации. В трудах зарубежных авторов касательно розничной торговли отражена информация о новых технологиях организации розничного бизнеса: Э. Голдратт, Т. Уоллас, Р. Сталь, F. Caro, J. Gallien, C. Crum, G. Palmatier, N. Lichtenstein, S. Tayur, R. Ganeshan, M. Magazine, D. Bartmann, M. F. Bach. Тем не менее, во многих из этих источников уделяется недостаточно много внимания важности прогнозирования спроса в розничном бизнесе, либо в его методологическом описании в качестве инструментов используются упрощенные экономико-математические модели, не учитывающие возможные нелинейные связи между факторами.

Прогнозирование социально-экономических и иных показателей, которые можно охарактеризовать таким общим понятием как временной ряд, описано в исследованиях ряда зарубежных и отечественных ученых: Р. Хиндмана, Дж. Атанасопулоса, J. S. Racine, Q. Li, T. Hayfield, C. A. Айвазяна, Б. Б. Демешева, И. С. Светунькова, С. Г. Светунькова, N. Kourentzes, А. И. Орлова, В. Н. Афанасьева, М. М. Юзбашева, О. А. Мишулиной, О. М. Писаревой, Н. А. Садовниковой, Р. А. Шмойловой, Т. А. Дубровой, В. К. Семёнычева, Е. В. Семёнычева. Исследования посвящены изучению динамики социально-экономических показателей, их свойствам и математическому моделированию процессов и временных рядов. Представлен широкий обзор тренд-сезонных и авторегрессионных инструментов, но практически не рассматривается сложное многоуровневое моделирование динамических процессов.

Темы продвинутого математического прогнозирования, статистического и машинного обучения, которые тесно связаны с задачами прогнозирования и оптимизации в розничной торговле проработаны в ведущих исследованиях современных зарубежных и отечественных авторов: Г. Джеймса, Д. Уиттона, Т. Хасты, Р. Тибширани, Э. Ына, Л. Бреймана, А. Мюллера, С. Гвидо, У. Маккинли,

С. Рашка, Я. Гудфеллоу, И. Бенджио, А. Курвилля, Дж. Хинтона, П. Флаха, Д. Кука, Л. П. Коэльо, П. Домингоса, С. Осовского, Р. Сичосз, Ю. И. Журавлева, Л. Н. Ясницкого, К. В. Воронцова, Д. П. Ветрова, А. В. Груздева, А. Б. Меркова. В источниках раскрыты основные достижения в области компьютерно-математического моделирования процессов и явлений тех процессов, которые описываются большим массивом учетных данных. Представленные примеры задач из практики ограничены классическими примерами статистического и машинного обучения: биологические задачи, задачи распознавания изображений, прогнозирование несложных процессов, т.е. отсутствует глубокое погружение в ту или иную область применения, в которой может быть использовано продвинутое математическое прогнозирование.

На текущий момент на отечественном рынке реализован ряд программных продуктов по прогнозированию спроса и управлению запасами такие как «Forecsys Goods4Cast», «Forecast NOW!», «Deductor», «Прогноз». На основе разработок и решений, реализованных с помощью рассмотренных продуктов, написано ряд статей таких авторов как Н. Б. Паклина, В. И. Орешкова, Ш. Аюбира, Бариновой О. В., А. А. Грицяя и многих других. Источники посвящены в основном методикам исследования данных, прогнозированию спроса и математическим алгоритмам, лежащим в основе программных комплексов по прогнозированию спроса. Большинство из рассмотренных программных приложений разработаны по концепции «черного ящика», когда пользователь не может влиять на структуру разработанных алгоритмов, что является проблемой для адаптации и масштабирования указанных программных продуктов.

Проблема разработки информативных моделей, апробации и внедрения программных решений в области прогнозирования спроса на предприятии розничной торговли определило выбор объекта, предмета, цели и задач диссертационного исследования.

Объектом исследования является предприятие розничной торговли как система товарных потоков и информации о них.

Предметом исследования является процесс товародвижения предприятия розничной торговли.

Целью диссертационного исследования является теоретическое и практическое развитие прогнозного моделирования покупательского спроса на предприятиях розничной торговли на основе методов машинного обучения.

Для достижения цели в данной работе необходимо решить следующие исследовательские и практико-ориентированные **задачи**:

1. Построить эмпирическую модель прогнозирования товарного спроса на основе данных пространственно-временной выборки.
2. Построить систему прогнозирования для переменных, которые включены как предикторы в основную модель спроса розничного предприятия.
3. Разработать программный комплекс для оценки будущего розничного спроса на основе построенной модели прогнозирования.

Область исследования соответствует паспорту научной специальности ВАК РФ 08.00.13 «Математические и инструментальные методы экономики» по следующим пунктам:

1.4. Разработка и исследование моделей и математических методов анализа микроэкономических процессов и систем: отраслей народного хозяйства, фирм и предприятий, домашних хозяйств, рынков, механизмов формирования спроса и потребления, способов количественной оценки предпринимательских рисков и обоснования инвестиционных решений.

2.3. Разработка систем поддержки принятия решений для рационализации организационных структур и оптимизации управления экономикой на всех уровнях.

Теоретическая и методологическая основа исследования

Теоретическую основу данной диссертационной работы составляют исследования в области торгового дела и экономико-математического прогнозирования. Для структурного анализа объекта исследования применяется инструментарий экономики торговых организаций с целью обозначить принципы функционирования торгового предприятия и область применения разрабатываемой экономико-математической модели. Используемый инструментарий для решения основной задачи – создания модели прогнозирования спроса – основан на классических и современных методах машинного обучения: классической линейной регрессии, линейной регрессии с регуляризацией, деревьях решений и их ансамблевых реализаций в виде случайного леса и бустинга.

Научная новизна. В ходе проведенного исследования получены результаты, которые обладают научной новизной и являются предметом защиты:

1. Построена модель прогнозирования спроса конкретной товарной позиции на основе пространственно-временной выборки данных с применением современных методов машинного обучения, позволяющая учитывать, в отличие от существующих регрессионных моделей прогнозирования спроса, особенности мультимодального спроса на товар (1.4. Разработка и исследование моделей и математических методов анализа микроэкономических процессов и систем: отраслей народного хозяйства, фирм и предприятий, домашних хозяйств, рынков, механизмов формирования спроса и потребления, способов количественной оценки предпринимательских рисков и обоснования инвестиционных решений. Глава 2, параграфы 2.5 и 2.6, стр. 61-70).
2. Построена оригинальная методика расчета будущих значений ключевых переменных математической модели прогнозирования спроса, увеличивающая прогностическую точность за счет использования современных инструментов анализа временных рядов: метод Prophet, байесовские временные ряды, современные версии алгоритмов ARIMA и экспоненциального сглаживания. (1.4. Разработка и исследование моделей и математических методов анализа микроэкономических процессов и систем: отраслей народного хозяйства, фирм и предприятий, домашних

хозяйств, рынков, механизмов формирования спроса и потребления, способов количественной оценки предпринимательских рисков и обоснования инвестиционных решений. Глава 2, параграф 2.4, стр. 52-61).

3. Разработан программный комплекс прогнозирования спроса на основе языка программирования R, функционирующий как сервис, встроенный в автоматическую систему заказа товара на предприятии розничной торговли, и способствующий удовлетворению покупательского спроса и оптимизации товарных запасов (2.3. Разработка систем поддержки принятия решений для рационализации организационных структур и оптимизации управления экономикой на всех уровнях. Глава 3, параграфы 3.4, 3.5, стр. 111-116).

Научная и практическая значимость результатов исследования

Научная значимость исследования состоит в критической оценке текущих методов и моделей прогнозирования товарного спроса и разработке новой экономико-математической модели, которая позволяет учитывать динамику спроса и экзогенных факторов. Здесь учитывается эффект многономенклатурности выбора товаров на итоговую потребность покупателя и эффекты, которые связаны с покупательским поведением, эластичностью спроса. Результаты, полученные в работе, вносят вклад в решение одной из самых важных народно-хозяйственных проблем повышения эффективности инструментов прогнозирования спроса и оптимизации управления товарными запасами в качестве приложения результатов моделирования. Практическая значимость работы заключается в реализации программного продукта в виде сервиса на языке R, который в автоматическом режиме осуществляет прогнозирования потребности покупателей в товаре. Алгоритм в основе работы сервиса является универсальным для торговых сетей, продающих товары повседневного спроса.

Апробация результатов исследования. Результаты диссертационного исследования были представлены на Международной научной конференции студентов, аспирантов и молодых учёных «Ломоносов-2017» (Москва, 2017 г.), Международной научно-технической конференции «Перспективные информационные технологии» (Самара, 2017 г.), Всероссийской заочной научно-практической конференции «Математические методы и интеллектуальные системы в экономике и образовании» (Ижевск, 2015 г., 2016 г., 2017 г.), Международной молодёжной научно-практической конференции «Математическое и компьютерное моделирование в экономике, страховании и управлении рисками» (Саратов, 2017 г.). Результаты исследования использовались при проведении занятий по дисциплинам «Моделирование бизнес-процессов», «Эконометрическое моделирование» и «Информационные системы управления производственной компанией» для студентов бакалавриата по направлению «Бизнес-информатика» Института экономики и управления Удмуртского государственного университета.

Результаты диссертационного исследования находятся на стадии активного внедрения в бизнес-процессы розничного предприятия ООО

«Гастроном». Применяя разработанные в диссертационной работе экономико-математические модели, ритейлер улучшил качество прогнозирования и разработки планов на ключевые показатели предприятия, увеличил уровень продаж по ряду товарных групп и оптимизировал товарный запас, что подтверждено Актом о внедрении результатов диссертационной работы.

Публикации. По теме диссертационного исследования опубликовано 10 работ объемом 7,11 п. л., из них в ведущих рецензируемых научных журналах и изданиях, определенных ВАК, - 4.

Структура и объем работы. Диссертация содержит введение, 3 главы и заключение, изложенные на 145 страницах машинописного текста. В работу включены 57 рисунков, 20 таблиц, 6 приложений и список литературы из 125 наименований.

Введение содержит описание актуальности темы, формулировку целей и задач работы, раскрывает основные научные методы, используемые в работе, основные положения, выносимые на защиту, и определяет содержательную часть работы.

В первой главе проведен анализ системы товародвижения на предприятии розничной торговли. Рассмотрено понятие покупательского спроса, его ключевые аспекты, а также важность задачи моделирования и прогнозирования. Сделан акцент на особенностях моделирования покупательского спроса в рамках настоящего диссертационного исследования. Приведены математические методы, которые рассматриваются как основные инструменты моделирования и прогнозирования спроса.

Во второй главе разработана методология прогнозирования ключевых переменных и прогнозирования целевой переменной – покупательского спроса. Модели прогнозирования ключевых показателей выстроены на классических и современных методах прогнозирования временных рядов. Разработанная модель прогнозирования спроса базируется на работе с панельными выборками, кластеризации временных рядов, обогащении исходных данных переменными экономического толка и применении методов машинного обучения. Приведенные модели оценены с помощью стандартных метрик качества для регрессий.

В третьей главе на основании данных, предоставленных ООО «Гастроном», рассчитаны оценки параметров и гиперпараметров приведенных регрессионных моделей. Оценена совмещенная модель из разных представленных методов на основе средневзвешенных оценок прогноза. Получены результаты по прогнозированию спроса и проведен сравнительный анализ с фактической ситуацией, на основе которого делаются выводы об экономической эффективности проведенной работы. Отражена архитектура и программная среда разработанной системы поддержки принятия решения, созданной на основе описанной методологии.

В заключении содержится описание основных выводов и результатов исследования.

II. ОСНОВНЫЕ ПОЛОЖЕНИЯ, ВЫНОСИМЫЕ НА ЗАЩИТУ

1. Построена модель прогнозирования спроса конкретной товарной позиции на основе пространственно-временной выборки данных, включающая авторские переменные, с применением современных методов машинного обучения, позволяющая учитывать, в отличие от существующих регрессионных моделей прогнозирования спроса, особенности мультимодального спроса на товар.

Большинство современных моделей и концепций исследования покупательского спроса основаны на малом количестве информации: учитываются только динамика самих значений спроса и некоторые ценовые факторы. Здесь под покупательским спросом понимается количественная потребность клиента розничной компании в товаре, обеспеченная его денежными средствами. В рамках диссертационного исследования целевой переменной u является спрос на товар, выраженный как складская учетная единица. Для решения проблемы использования малоинформативных моделей была сформирована пространственно-временная (панельная) выборка с такими переменными как лаговые переменные спроса; цена товара; наличие промо-акций; календарные признаки: принадлежность к дню недели, праздничному или предпраздничному периоду; наименование производителя и страны-производителя товара; геометрические характеристики товара; температурный режим в зоне действия торговой сети; суммарные статистики по количеству чеков; количеству продаж товарной группы, к которой относится целевой товар; количество товарных позиций, взаимозаменяемых по цене.

Приведем алгоритмы построения двух оригинальных переменных, которые также участвуют в моделировании: количество товарных позиций, взаимозаменяемых по цене и товарные кластеры. Переменные «Количество товарных позиций, взаимозаменяемых по цене» и «Товарные кластеры» являются авторской разработкой, которые добавляют существенно новую информацию об объекте моделирования – спросе, а также позволяют использовать преимущества панельной структуры данных, используя информацию обо всем историческом спросе номенклатуры товаров.

Для того, чтобы рассчитать количество взаимозаменяемых позиций N_i было определено k ценовых интервалов $(pr_i, pr_i + h]$, где pr_i – цена товара (нижняя граница i -ого интервала), h – шаг цены, формируемый исходя из следующего принципа:

$$h = (pr_\alpha - pr_{min}) \times d, \quad (1)$$

где pr_α – α -квантиль распределения цен группы товаров (в рассматриваемой задаче вероятность α принимается равной 0,9), pr_{min} – минимальное значение цены по группе товаров, d – коэффициент максимально возможного изменения взаимозаменяемых цен, выраженный в процентах. Значение pr_α использовалось при определении шага для объединения экстремально высоких значений цен в группу « pr_α и выше», т.е. для определения $k + 1$ ценового диапазона. Количество товарных позиций, взаимозаменяемых по цене N_{SKU} – это количество складских учетных единиц, принадлежащих к одному ценовому

диапазону с товаром, спрос которого моделировался: если $pr_{SKU} \in (pr_i, pr_i + h]$, то $N_{SKU} = N_i$, где pr_{SKU} – значение стоимости складской учетной единицы. Исходя из формулы (1), значение показателя N_{SKU} зависит от ширины ценового диапазона и параметра d . Значение параметра d было подобрано по децильному принципу: все номенклатурные единицы разделились на ψ равнозначных групп по объему.

Показатель N_{SKU} вводился в систему переменных для отражения предположения о том, что товары с ценой в рамках одного диапазона имеют показатель средних продаж $\bar{s}_i = S_i/N_i \rightarrow 0$ при $N_i \rightarrow +\infty$, где S_i – суммарное количество продаж товаров в рамках одного ценового диапазона $(pr_i, pr_i + h]$, N_i – количество товарных позиций внутри заданного ценового диапазона, \bar{s}_i – средние продажи товаров на 1 наименование внутри i ценового диапазона.

Так как для прогнозирования спроса использовалась выборка с большим количеством товаров, то возникла проблема потери информации при разбиении данных на достаточно мелкие выборки. В случае этого разбиения, моделирование целевой переменной могло происходить на малых данных, что привело бы к большой ошибке на тестовой выборке. В ходе реализации задачи прогнозирования спроса для ежедневного потребления набора товаров были выявлены следующие характеристики:

1. Товарный набор $Y = (y_1, \dots, y_i, \dots, y_N)$ не является однородным по времени, поэтому при прогнозировании были использованы временные ряды спроса каждого товара $y_i = (y_i(1), y_i(2), \dots, y_i(t_i))$, где t_i – момент времени измерения спроса i -ого товара.
2. Информация о двух товарах из набора может быть не равнозначной по времени обзора. Если товар со спросом y_j является новинкой, а значения y_i относятся к товару со зрелой стадией жизненного цикла, то длина временного ряда y_i будет существенно больше y_j : $t_i \gg t_j$.
3. Набор независимых величин X , присущих как набору товаров, так и каждому товару в отдельности, также может меняться во времени: $X = (x(1), x(2), \dots, x(t_i))$.

Исходя из указанного следует, что при достаточно большом количестве N элементов в наборе, неравномерно распределенных на всем временном промежутке T решения регрессионной задачи, качество функционала для каждой случайно взятой оценки спроса y_i начинает падать. Одним из способов преодоления этого ограничения является применение кластерного анализа к набору товаров с целью выделить определенные подгруппы похожих товаров внутри набора. Для проведения кластерного анализа были использованы два типа данных о товарах в наборе:

- Данные об измерениях временных рядов по величинам спроса в наборе Y , т.е. непосредственно кластеризуемые временные ряды.
- Данные о свойствах товаров в наборе. К ним относится информация о стоимости товара, его производителе, геометрические характеристики.

Кластерное разбиение производилось на данных по торгово-розничной сети города Ижевска. Выделение кластеров внутри набора осуществлялось с помощью классических методов кластерного анализа: алгоритма k-средних и EM-алгоритма (выделение гауссовых смесей распределения). Выбор данных алгоритмов был обусловлен именно простотой исполнения и улучшением качества решения задачи прогнозирования. Алгоритм кластерного анализа был построен на многих переменных, характеризующих товарную позицию (таблица 1).

Таблица 1 – Набор переменных для кластеризации

Наименование переменной	Описание переменной
<i>weekday</i>	День недели даты, на которую осуществляется прогноз <i>i</i> -ого товара.
<i>Y1...Y28</i>	Значения спроса y^i для каждого товара в наборе за последние 28 дней (не включая прогнозируемую величину). Над исходными величинами была произведена процедура $\log(y + 1)$
<i>Max, Min, Median, Mean, Sd</i>	Основные структурные показатели по спросу за последний период 28 дней продаж – максимальный, минимальный, медианный и средний спрос соответственно, а также стандартное отклонение спроса.
<i>lin</i>	Структурный показатель, который выражает коэффициент линейного тренда по спросу за последние 28 дней. Рассчитывается с помощью метода наименьших квадратов.
<i>numberNull</i>	Структурный показатель, который выражает количество нулевых значений спроса за последние 28 дней
<i>P1...P28</i>	Наличие / отсутствие промо-акции на товар за последние 28 дней.
<i>price</i>	Стоимость за 1 единицу товара.
<i>weight</i>	Вес товара. Здесь, вместо веса, могут быть использованы любые важные геометрические характеристики товара.
<i>country</i>	Страна-производитель продукта. Данные включаются в виде набора фиктивных переменных.
<i>maker</i>	Наименование производителя продукта. Данные включаются в виде набора фиктивных переменных.

Ключевой метрикой оценки эффективности кластерного анализа является оценка среднеквадратического ошибки *MSE* итогового линейного прогноза спроса на тестовой выборке. Для построения кластеров использовались данные по одному из магазинов сети «Гастроном» за период 01.01.2013 по 30.09.2016 со всеми переменными, перечисленными в таблице 2. Среднеквадратическая ошибка *MSE* рассчитывалась строго на тестовой выборке, которая определена правилом деления 70 на 30 (70% попадает в обучение, 30% - в тест). Предварительно был оценен размер *MSE* для линейной модели без выделения кластеров, который равен 9,369712. По итогам реализации компьютерных экспериментов по выделению кластеров по двум алгоритмам были рассчитаны оценки ключевой метрики (таблица 2).

Таблица 2 – Значение MSE на тестовой выборке при реализации кластерного анализа

Вид алгоритма	Количество кластеров			
	2	5	10	15
K-means	8,88359	8,93621	8,94339	8,95204

EM-алгоритм	8,97449	8,87694	9,96271	8,87094
-------------	---------	---------	---------	---------

Качество прогнозирования значительно увеличивается при EM-алгоритме, что позволило сделать вывод об его использовании при реализации основной задачи. В качестве готового решения был использован EM-алгоритм с 15 кластерами. На основе данного разбиения кластеров, выведем минимальный MSE равный 8,87094, что на 5,3% лучше, чем модель без использования кластерного разбиения. Кластерная структура позволила свести некоторые оценки влияния независимых факторов X на Y с учетом наличия в наборе товаров разного рода подгрупп.

На основе панельной выборки из указанных переменных была построена математическая модель прогнозирования спроса. Иерархия моделирования представлена на рисунке 1 и является системой прогнозных моделей, которые выстраиваются для каждой товарной группы каждого магазина розничной сети.

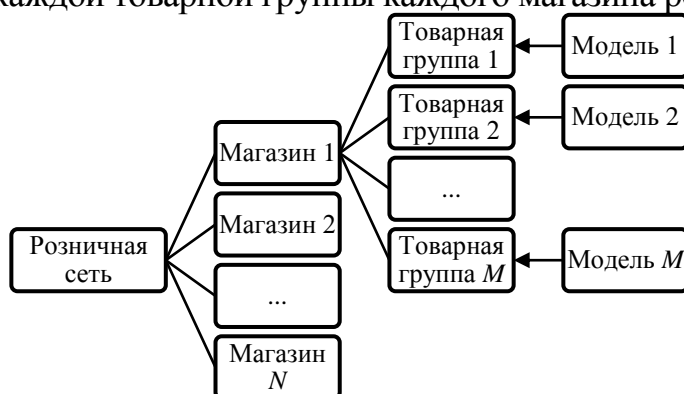


Рисунок 1 – Иерархия прогнозного моделирования

В ходе статистических исследований было обнаружено большое количество товаров с неравномерно распределенным ежедневным спросом. В теории это означает существование бимодального (мультимодального) распределения спроса на товар. Первая мода при этом находится в нуле (рисунок 2).

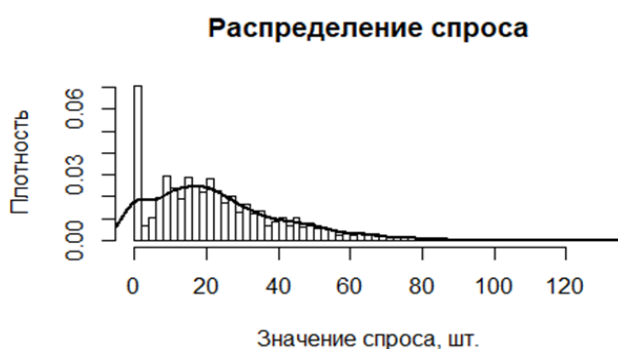


Рисунок 2 - Распределение спроса на товар

Подобная структура накладывает ограничение на прямое применение регрессионного подхода при прогнозировании. Поэтому задача прогнозирования спроса на товар была представлена следующими этапами:

1. Определение вероятности ненулевого спроса на товар $P(y \neq 0)$. Расчет сводится к решению задачи классификации для установленного панельного набора данных;

2. Расчет прогнозного значения для всей совокупности случаев ненулевого спроса:

$$\hat{D} = f(x|y \neq 0), \quad (2)$$

где \hat{D} – оценка спроса (регрессионная). Здесь речь идет о решении задачи регрессии на выборке с отсутствием значений нулевого спроса.

3. Подведение итоговой оценки прогнозного значения спроса как математического ожидания спроса на товар:

$$\hat{y} = E(y) = P(y \neq 0) \cdot \hat{D} \quad (3)$$

Численные эксперименты показали, что указанный подход является оправданным с точки зрения оптимизации целевых метрик. На финальном этапе после моделирования с помощью методов машинного обучения была произведена комбинация результатов (рисунок 3).

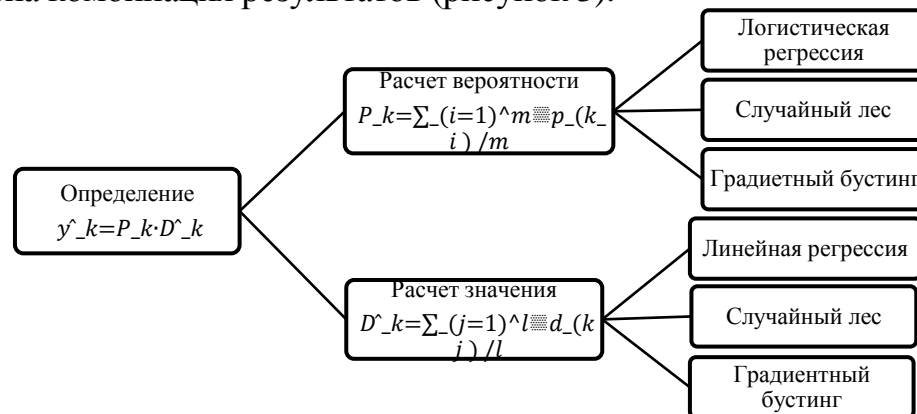


Рисунок 3 – Система моделирования товарного спроса

По рисунку видно, что комбинация результатов осуществляется простой процедурой среднего арифметического по всем m методам для классификации или определения вероятности ненулевого спроса и по всем l регрессионным методам. По формуле (3) итоговое решение выглядит следующим образом:

$$\hat{y} = E(y_k) = P_k(y \neq 0) \cdot \hat{D}_k = \sum_{i=1}^m p_{k_i} / m \cdot \sum_{j=1}^l d_{k_j} / l \quad (4)$$

где p_{k_i} – оценка вероятности ненулевого спроса, рассчитанные i -м методом, d_{k_j} – оценка регрессионного значения спроса, рассчитанного j -м методом.

Рассмотрим усреднение для оценки вероятности ненулевого спроса:

$$P_k(y \neq 0) = \sum_{i=1}^m p_{k_i} / m = \sum_{i=1}^3 p_{k_i} / 3, \quad (5)$$

где $m = 3$ ввиду использования 3-х моделей оценки вероятности, потому как выбираются методы логистической регрессии, случайного леса и градиентного бустинга. Это сделано ввиду того, что значение корреляции между результатами логистической регрессии и логистической регрессии с регуляризацией высоко, а качество регрессии с регуляризацией хуже.

Приведем характеристики моделей, составляющий ансамбль. Логистическая регрессия имеет следующую спецификацию:

$$p(X) = \frac{e^{BX}}{1 + e^{BX}},$$

$$BX = \beta_0 + (a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + a_5 y_{t-5} + a_6 y_{t-6}) \times B_1 C + B_2 X_2, \quad (6)$$

Здесь, β_0 – смещение, a_1, a_2, a_3, a_5, a_6 и $y_{t-1}, y_{t-2}, y_{t-3}, y_{t-5}, y_{t-6}$ – коэффициенты и значение лагов спроса соответственно, B_1 – набор коэффициентов и C – фиктивный набор переменных, которые отражают принадлежность объекта прогнозирования к товарному кластеру, X_2 – набор прочих предикторов с матрицей коэффициентов B_2 . Полный набор предикторов, использованных в моделировании, состоит из 122 переменных. Значение $AUC = 0.844$ принимает на тестовой выборке.

Случайный лес классификации использует те же переменные при моделировании. Здесь и далее, гиперпараметры алгоритмов машинного обучения подбирались с помощью метода перекрестной проверки на заранее заданной сетке параметров. Параметры случайного леса классификации были признаны оптимальными на следующих значениях: количество деревьев $N = 500$, количество признаков, выбираемых случайно для каждого построения дерева $m = 5$ и количество объектов в узлах формируемых деревьев $MN = 200$. В данном случае оценка $AUC = 0.853$ на тестовой выборке.

Градиентный бустинг обладает большим списком настраиваемых параметров (таблица 3).

Таблица 3 – Оптимальные значения параметров градиентного бустинга для задач классификации и регрессии

Обозначение параметра	Описание параметра	Значение параметра для классификации	Значение параметра для регрессии
B	Тип базового алгоритма для бустинга	Дерево решений	Дерево решений
η	Скорость обучения алгоритма	0.1	0.1
N	Число итераций алгоритма	300	200
MD	Максимальная глубина деревьев	6	6
MC	Минимальное количество объектов в листе дерева	5	10
δ	Доля объектов выборки, используемые на каждой итерации алгоритма	1	0.9
c	Доля используемых признаков для каждой итерации	0.5	0.7

Далее выводится ROC-кривая по качеству ансамбля на тестовой выборке в сравнении с ROC-кривой с градиентным бустингом как с лучшим методом по качеству из моделируемых (рис. 4).

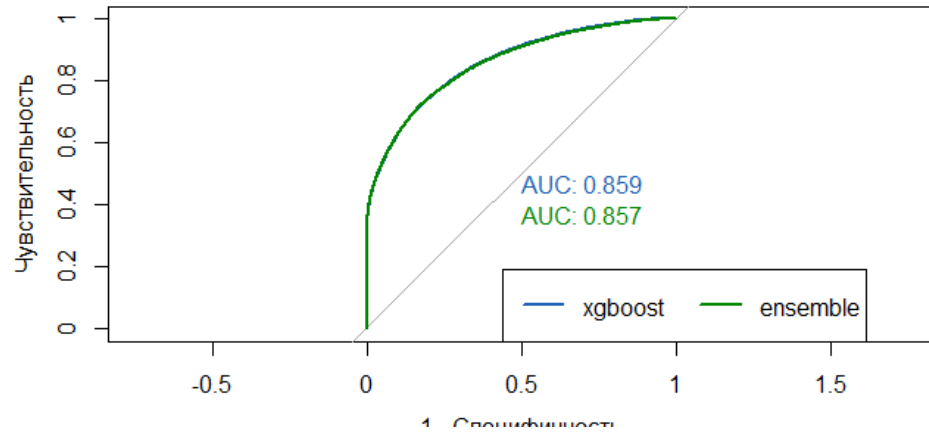


Рисунок 4 – ROC-кривые по лучшему методу и комбинации

Видно, что ансамбль не превзошел по качеству лучший метод. При этом, при переборе других комбинаций также не была достигнута оценка $AUC > 0.859$. Поэтому для итоговой оценки, первоначально, выбирается результат по методу градиентного бустинга.

Для регрессионной оценки \hat{D} также выводится ансамбль прогнозов по принципу:

$$\hat{D}_k = \sum_{j=1}^l d_{k_j} / l = \sum_{j=1}^3 d_{k_j} / 3 \quad (7)$$

Количество методов $l = 3$ ввиду того, что, аналогично предыдущей ситуации, прогнозы по линейной регрессии и линейной регрессии с регуляризацией сильно коррелированы. В данном случае в качестве базового линейного метода был выбран метод с регуляризацией, так как он работает точнее. Прогнозное решение на основе линейной регрессии имеет структуру:

$$D = \beta_0 + (a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + a_5 y_{t-5} + a_6 y_{t-6}) \cdot (B_1 C + B_{21} X_2) + B_{22} X_2 + B_3 X_3, \quad (8)$$

где X_2 – часть предикторов, связанных с целевым значением мультипликативно, X_3 – независимые переменные, связанные со спросом аддитивно. Количество переменных в модели равно 208. Оптимальные гиперпараметры регуляризации $\alpha = 0,25$ и $\lambda = 0.01$, где α – сочетание штрафов L1 и L2, λ – параметр, отвечающий за уровень сжатия коэффициентов регрессии. Оценка качества полученной модели характеризуется значениями следующих метрик качества на тестовой выборке: $MSE_{ts} = 23.50026$ $MAE_{ts} = -0.09650419$. Регрессионный случайный лес имеет следующие оптимальные значения параметров: $N = 500$, $m = 12$ и $MN = 10$. Значение метрик на тестовой выборке соответствуют $MSE_{ts} = 22.33565$ и $MAE_{ts} = 0.1398928$. Оптимальные значения параметров для градиентного бустинга над регрессионными деревьями приведены в таблице 8. Качество метода на тестовой выборке оценено $MSE_{ts} = 22.16758$ и $MAE_{ts} = -0.130474$.

После расчета средней по 3 методам проведем оценку значений метрик на тестовой выборке:

$$MSE_{ts} = 21.77471 \quad MAE_{ts} = -0.04042693 \quad (9)$$

Соответственно, для расчета итогового значения прогноза спроса совместили оценки вероятности ненулевого спроса по бустингу и комбинации регрессионных оценок по формуле (8). Результат итоговой метрики на тестовой выборке следующим образом:

$$MSE_{ts} = 7.990168 \quad MAE_{ts} = -0.04111812 \quad (10)$$

В данном случае под тестовой выборкой понимается полный объем информации, где $D \geq 0$. Кроме того, в ходе дополнительных экспериментов были получены более оптимальные значения MSE и MAE :

$$MSE_{ts} = 7.941281 \quad MAE_{ts} = -0.0311855 \quad (11)$$

Данная оценка была получена при замене метода прогнозирования вероятности ненулевого спроса с градиентного бустинга на алгоритм логистической регрессии. Это является более простой и устойчивой оценкой.

Согласно формуле (4) итоговая математическая модель прогнозирования спроса выглядит следующим образом:

$$\hat{y} = p(X) \cdot \left(d_{k_{lin}}(X) + d_{k_{rf}}(X) + d_{k_{gbm}}(X) \right) / 3 \quad (12)$$

где $p(X)$ – модель вероятности ненулевого спроса на основе логистической регрессии, $d_{k_{lin}}(X)$, $d_{k_{rf}}(X)$ и $d_{k_{gbm}}(X)$ – модели регрессионной оценки спроса на методов основе линейной регрессии с регуляризацией, случайного леса и градиентного бустинга соответственно.

При реализации регрессионного моделирования ненулевого спроса \hat{D} были также сделаны выводы об указанном выше предположении, которое легло в основу смысла переменной «Количество товарных позиций, взаимозаменяемых по цене» N_{SKU} . Линейная модель \hat{D} была оценена с двумя вариантами переменной N_{SKU} : рассчитанной для всех товарных позиций и рассчитанной для всех товарных позиций с действующей промо-акцией (снижение цены) на момент прогнозирования по отношению к товару. Коэффициенты полученной модели линейной регрессии равны -0,0077 и -0,0071 (при среднедневном спросе по обучающей выборке 5,96 шт.; медианном спросе – 2 шт.). Данный результат можно интерпретировать следующим образом: при увеличении количества N_{SKU} на 1 единицу в рамках ценового диапазона $(pr_i, pr_i + h]$, в котором находится товар на который прогнозируется спрос, оценка спроса \hat{D} снижается на 0,0077 единиц (на 0,0071 в случае акционного товарного соседства). В условиях, когда в рамках одного ценового диапазона может продаваться не менее 20-30 товарных позиций это может привести к 2,5% - 4% снижению спроса на товар в среднем. Это позволяет в дальнейшем выводить рекомендации к ассортиментной политике на розничном предприятии.

По итогам моделирования сделан вывод, что модель эффективнее описывает спрос, чем большинство современных аналогов, основанных на тренд-сезонном анализе временного ряда. Это преимущество достигается с помощью включения в модель оригинальных переменных, имеющих высокоинформативные значения, использования панельной структуры данных (пространство товаров и времени их продажи), разделения исходного спроса на

компоненты – вероятности покупки $P_k (y \neq 0)$ и значения величины спроса \hat{D}_k – и моделирования их отдельно, использования в основе прогнозирования предикторов современной методики прогнозирования временных рядов, а также применения современных методов машинного обучения для построения итоговой модели.

2. Построена оригинальная методика расчета будущих значений ключевых переменных математической модели прогнозирования спроса, увеличивающая прогностическую точность за счет использования современных инструментов анализа временных рядов: метод Prophet, байесовские временные ряды, современные версии алгоритмов ARIMA и экспоненциального сглаживания.

Для ряда ключевых переменных – количества проданных товаров, средней температуры и количество чеков, которые включены в модель прогнозирования покупательского спроса по каждому отдельному товару, необходимо было дать собственный прогноз. Это потребовалось, чтобы избежать использования значений переменной из будущего или так называемого явления «протечки в данных» (от англ. data leakage). Следовательно, была разработана методика прогнозирования указанных переменных на основе инструментов анализа временных рядов.

В основе прогнозирования лежат следующие методы:

1. Модель экспоненциального сглаживания (ETS).
2. Модель авторегрессии – скользящего среднего (ARIMA).
3. Модель комплекснозначного экспоненциального сглаживания (CES).
4. Модель Prophet (технология от Facebook).
5. Байесовская структурная модель временных рядов (BSTS).

Указанные методы прогнозирования временных рядов используются для моделирования трех ключевых переменных, которые имеют потенциальный эффект на спрос: иерархически связанный со спросом параметр – количество продаж в товарной группе; параметр, определяющий емкость спроса в рамках розничной точки – количество чеков; параметр, который имеет сильное влияние на товары повседневного спроса – температурный режим. Линейные модели, которые легли в основу прогноза спроса, позволяют оценить влияние указанных показателей на покупательский спрос. Это видно по Таблице 4.

Таблица 4 – Коэффициенты линейных моделей покупательского спроса для ключевых переменных

Полное наименование ключевых переменных	К-ты логистической модели	Связанная переменная (мультипликативно)	К-ты линейной модели
Количество продаж в товарной группе	0,000720	Лар L1	0,000306
		Лар L2	0,000099
		Лар L3	0,000070
		Лар L4	0,000198
		Лар L5	0,000160
		Лар L6	-0,000028

		Лар L7	-0,000062
Количество чеков	0,000029	-	-0,000105
Температурный режим (среднедневная температура)	-0,001557	-	-0,004793

Согласно таблице 4 видно, что с ростом количества продаж в товарной группе растет и вероятность ненулевого спроса, а связь этого показателя с величиной спроса определяется через мультипликативное отношение с лагами спроса. Например, при значении количества продаж равном 1000 логарифм шансов ненулевого спроса увеличивается в 0.72 раза, а сама величина спроса увеличивается на 0,306 от первого лага в штуках (не берем во внимание остальные лаги от значений спроса). Это показывает четкую положительную взаимосвязь между общим масштабом продаж и значениями индивидуального спроса. Интересную интерпретацию имеет взаимосвязь показателя количества чеков и оценки спроса: при росте показателя растет и вероятность покупки, но при этом падает сама величина спроса. Это характерно для данного вида товаров, когда при большом наплыве покупателей спрос стабилизируется, но при этом он замещается другими продуктами, имеющими схожие потребительские характеристики. Рост среднедневной температуры ведет как к снижению вероятности покупки, так и снижению величины спроса: на каждый 1 градус уменьшение логарифма шансов в 0,001557 раз, а величина уменьшается на 0,004793 штуку. Стоит отметить, что с учетом сложности модели, данная интерпретация не является исчерпывающей, но позволяет судить о важности указанных переменных.

Для прогнозирования ключевых переменных принято решение выбрать за основную метрику качества корень от среднеквадратической ошибки модели (*RMSE*). В качестве дополнительной метрики для оценки качества была определена средняя абсолютная ошибка в процентах (*MAPE*). Данная система метрик позволила объективно оценить точность моделей на тестовой выборке и сравнить используемые методы. Прогнозирование количества проданных товаров производилось с помощью данных по продажам одной из товарных групп розничной точки сети «Гастроном». Период выборки для всех прогнозируемых величин: 01.05.2009 по 30.09.2016. Временной ряд разбивается на обучающий – с 01.05.2009 по 31.12.2015 – и на тестовый – с 01.01.2016 по 30.09.2016.

Основные метрики по методам прогнозов продаж по группе представлены в таблице 5.

Таблица 5 – Результат прогнозирования для прогноза продаж

Метрика качества	Prophet	ARIMA	ETS	CES	BSTS	Mean (ARIMA, ETS)
<i>RMSE</i>	170,21	136,66	156,00	195,06	160,37	142,96
<i>MAPE</i> , %	28,93	18,52	20,40	28,59	22,89	18,76

По оценке основной метрики *RMSE* комбинированный прогноз из лучших методов менее оптимален, поэтому было принято решение об использовании *ARIMA* в качестве итогового. Точечные экспертные корректировки в

праздничные дни и при реализации сложных маркетинговых мероприятий позволят улучшить качество прогноза. Итоговая динамика прогнозирования представлена на рисунке 5.

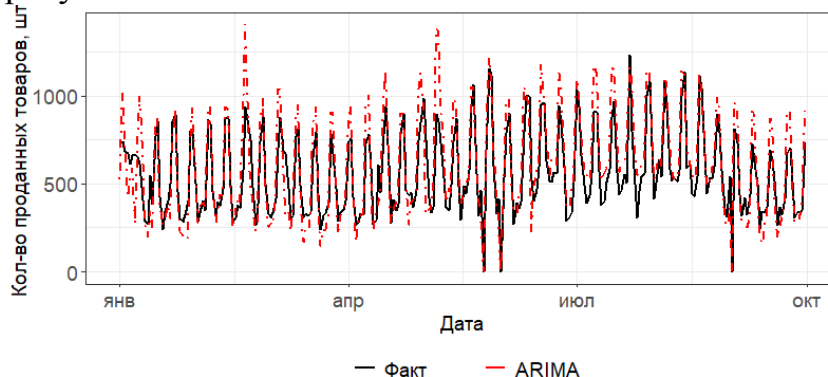


Рисунок 5 – Динамика продаж товарной группы с итоговым прогнозом

При исследовании влияния временного ряда температур на покупательский спрос было принято решение о необходимости в прогнозировании показателя. Здесь было выявлено несколько характеристик:

1. Видна четкая годовая сезонность – это является главной компонентой при прогнозировании.
2. Ряд имеет сигмоидальную структуру, что позволило успешно использовать ряды Фурье в качестве дополнительных предикторов.

На рисунке 6 приведен график прогнозов на тестовой выборке по трем методам прогнозирования, средний прогноз по двум лучшим методам и фактические значения температур.

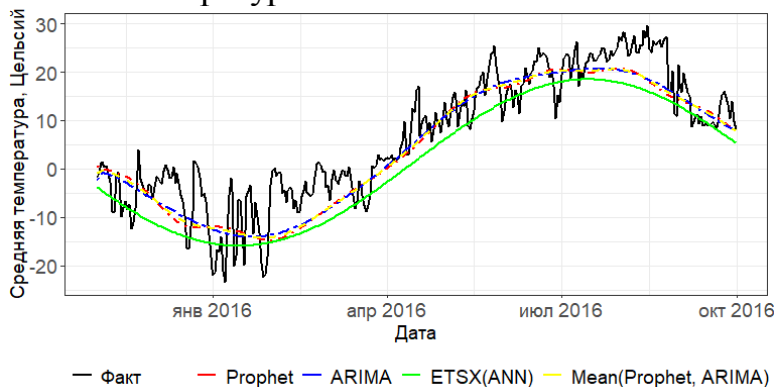


Рисунок 6 – График средних температур с прогнозами

Выбор только трех методов прогнозирования из представленных, обусловлен избыточностью применения более продвинутых методов прогнозирования погодных условий при малом объеме данных. В таблице 6 приведены с основные метрики прогнозов.

Таблица 6 – Результат прогнозирования для прогноза температуры

Метрика качества	Prophet	ARIMA	ETSX(ANN)	Mean(Prophet, ARIMA)
<i>RMSE</i>	5,60	5,36	6,85	5,46
<i>MAPE</i> , %	201,52	195,49	275,41	197,95

Метрика *MAPE* показала себя не лучшим образом, так как исходный временной ряд имеет как положительные, так и отрицательные значения.

Поэтому при оценке модели как основу рассматривалась $RMSE$, которая дает более четкую картину по методу. Наиболее эффективным алгоритмом является модель $ARIMA$.

Для прогнозирования показателя количества чеков использовались все рассматриваемые в работе методы. В основе моделирования находятся следующие переменные:

1. Данные о наличии государственного, негосударственного и религиозного праздника за период.
2. Сезонность по годовому и недельному периодам. Задавалась с помощью фиктивных переменных, а также рядов Фурье.
3. Месячная сезонность роста доходов населения (получение заработной платы). Подобная зависимость моделировалась полиномом 5-й степени от номера дня в месяце, что также включается практически во все модели кроме Prophet.

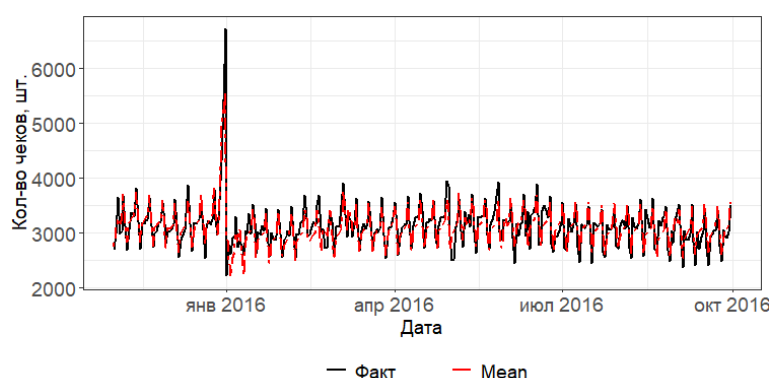


Рисунок 7 – График количества чеков со средним прогнозом

На рисунке 7 итоговый средний прогноз сформировался по результатам всех методов кроме модели экспоненциального сглаживания (ETS). Такой подход применялся из-за высокой коррелированности прогнозов по методам ETS и CES. При этом, по указанным метрикам лучшей из двух моделей является CES. В таблице 7 приведены метрики прогнозов.

Таблица 7 – Результат прогнозирования для количества чеков

Метрика качества	Prophet	ARIMA	ETSX(MNN)	CESX(n)	BSTS	Mean
$RMSE$	240,46	224,40	228,54	227,54	230,86	188,69
$MAPE$, %	6,05	5,26	5,30	5,40	5,47	4,43

Видно, что значение среднего арифметического по прогнозам имеет лучшие метрики в сравнении с остальными прогнозами.

Исходя из приведенных результатов вычислений, в качестве итогового решения целесообразно применять сочетания методов прогнозирования на основе усреднения результатов для увеличения качества прогнозирования. Выбор перечня методов, которые участвуют в ансамбле, обусловлен минимизацией основной и дополнительной метрик – $RMSE$ и $MAPE$.

Итоговое представление методики обозначено в схеме на рисунке 8.

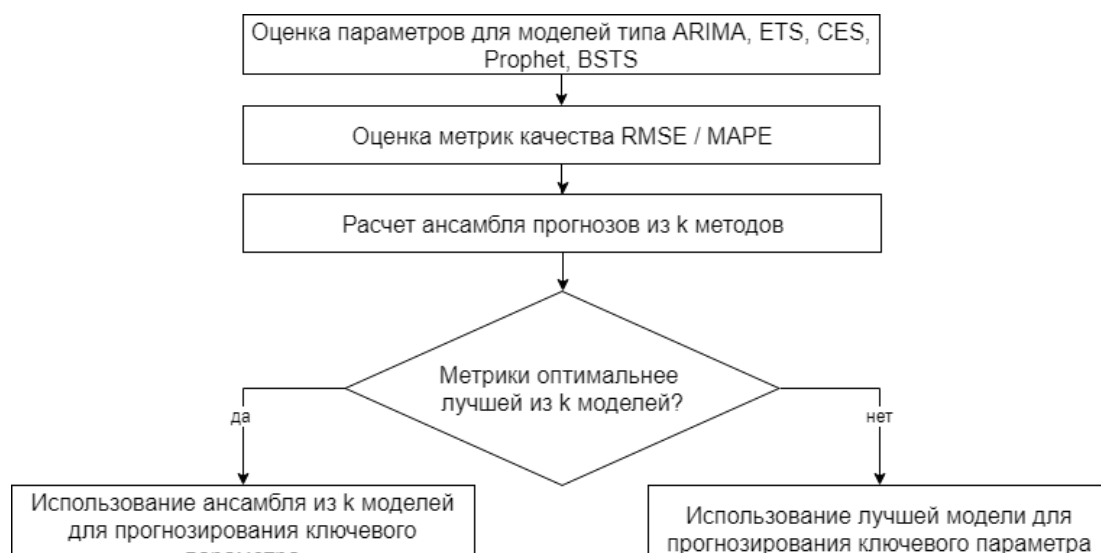


Рисунок 8 – Блок-схема методики прогнозирования ключевых параметров

Предлагаемая методика прогнозирования временных рядов позволила выбрать лучший метод или ансамбль моделей для оценки будущих значений переменных. При этом указанные переменные имеют значимое влияние на процесс принятия решений покупателем, что видно по интерпретации линейных коэффициентов. Например, правильная оценка роста количества продаж в группе позволит зафиксировать рост индивидуальной оценки спроса, что приведет к правильной оценке покупательской потребности. В целом, модели временных рядов для указанных переменных позволяют увеличить точность при принятии решений и планировании мероприятий. Методика регулирует процесс создания моделей временных рядов и их селекцию по уровню качества, которые затем используются в создании модели прогнозирования спроса и для участия в разработке маркетинговых планов розничного предприятия (количество продаж, чеков).

3. Разработан программный комплекс прогнозирования спроса на основе языка программирования R, функционирующий как сервис, встроенный в автоматическую систему заказа товара на предприятии розничной торговли, и способствующий удовлетворению покупательского спроса и оптимизации товарных запасов.

Язык R – мультипарадигмальный язык программирования для статистической обработки данных и визуализации. Язык R активно используется для исследований в академической среде и в последнее время получает все большую распространенность для решения бизнес-задач. Особенностью языка является удобство написания высокоуровневых решений на его основе и наличие функциональных расширений – пакетов.

Система прогнозирования – это прежде всего система обработки и анализа данных, на выходе которой исследователь имеет результат в виде модели прогнозирования и предсказания целевой переменной. Подобный класс задач решает такая область методов как Data Mining. В данном контексте этапы разработанной системы прогнозирования отражены на рисунке 9.

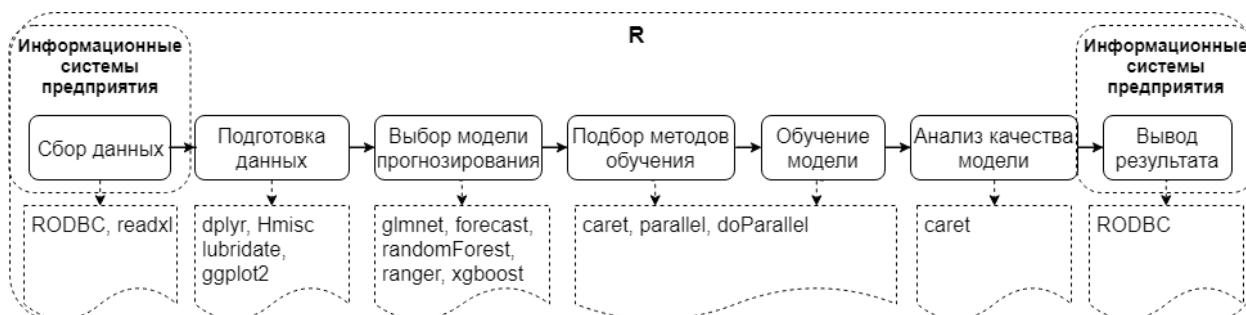


Рисунок 9 – Архитектура системы прогнозирования на языке R (с указанием доп. пакетов)

Как видно, возможности языка R охватывают все этапы работы системы прогнозирования. При этом, при подготовке решения не обошлось без использования пакетных расширений, широко распространённых среди разработчиков в среде R. Условно их можно разделить на 4 группы:

- Пакеты, которые осуществляют функции интеграции для сбора данных: *RODBC* (для интеграции с SQL-базами данных), *readxl* (для сбора данных с внешних Excel-файлов).
- Пакеты, которые осуществляют функции обработки и подготовки данных к моделированию: *dplyr* (облегчает фильтрацию, агрегацию и иные манипуляции с данными), *lubridate* (облегчает работу с временными данными), *Hmisc* (обработка пропущенных значений), *ggplot2* (позволяет визуализировать необходимые связи в данных).
- Пакеты, которые отвечают за создание и оценку качества моделирования: *glmnet* (линейные модели с регуляризацией), *forecast* (модели временных рядов), *randomForest* (реализация метода «случайный лес»), *ranger* (реализация вероятностного и классического методов «случайный лес»), *e1071* (реализация нескольких методов, в частности машины опорных векторов), *xgboost* (реализация градиентного бустинга), *caret* (отвечает за организацию моделирования и оценку качества результата);
- Пакеты, которые отвечают за повышение эффективности реализуемого программного комплекса: *parallel*, *doParallel* (позволяет ускорить процедуры за счет параллелизации процессов).

Полученная система прогнозирования спроса функционально интегрируется с глобальной информационной системой предприятия и используется в автоматической системе заказа товара. По сути, модуль на R – это микросервис, который интегрируется с информационной системой предприятия. Разработанное ПО используется в бизнес-процессах сети ООО «Гастроном», а также может быть использовано в розничных сетях схожего формата во всех регионах России. Это делает продукт исследования конкурентоспособным в рамках практической плоскости применения результатов диссертационного исследования.

Система прогнозирования спроса в приложении к автоматическому заказу товара на розничном предприятии позволила оптимизировать складские запасы розничного предприятия (таблица 8).

Таблица 8 – Показатели уровня товарных запасов (количество)

Показатель	Фактическое значение	Значение при применении прогноза	Прирост
Среднее значение товарных запасов, в шт.	53,99	13,57	-74,87%
Уровень товарных запасов в днях (среднее значение), в дн.	44,83	10,07	-77,54%
Показатель товарооборачиваемости (среднее значение), количество раз	0,03	0,13	290,56%

Характеристики сформированы на количественных измерениях: количество продаж и количество остатков. Внедрение системы прогнозирования при еженедельной поставке снижает уровень товарных запасов, увеличивает оборачиваемость товара, при этом продажи указанного товара остаются неизменными.

III. ЗАКЛЮЧЕНИЕ

В диссертационной работе построены экономико-математические модели прогнозирования ключевых показателей розничной торговли и потребительского спроса, которые снижают издержки предприятия:

1. Разработана методология построения математических моделей прогнозирования ключевых показателей, связанных с основными бизнес-процессами розничной торговли, основанная на инструментах анализа временных рядов. В контексте самостоятельного использования данная методология позволяет прогнозировать большое количество тренд-сезонных процессов в розничной торговле, а также может быть использована при разработке планов для показателей.
2. Разработана модель прогнозирования покупательского спроса на основе методов машинного обучения и показателей, характеризующих поведение покупателя. Предложенная модель используется в основе автоматизированного заказа товара, тем самым минимизируя издержки на обеспечение товарного запаса на предприятии.
3. Разработан программный комплекс на основе языка R, который в виде сервиса работает совместно с системами поддержки принятия решений на предприятиях сферы розничной торговли. В ходе разработки сервиса использовались расширенные возможности R, основанные на подключении дополнительных пакетов.
4. Разработанный комплекс математических моделей в виде сервиса (набора сервисов), реализованных на языке R, готов к внедрению на любых предприятиях розничной торговли, которые имеют достаточную историю продаж.
5. Применение разработанного сервиса в рамках компании ООО «Гастроном» позволило снизить средние товарные запасы на 9,6%, при этом увеличив продажи товара на 4,3%.

IV. СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в изданиях, входящих в перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук

1. Пивкин, К. С. Корреляционный анализ факторов влияния на покупательский спрос розничного магазина как этап формирования модели прогнозирования и управления запасами // Вестник Удмуртского университета. Серия Экономика и право. 2016. №3. С.40-50.
2. Пивкин, К. С. Алгоритм построения линейной модели на панельных данных как этап эконометрического прогнозирования товарного спроса // Вестник Удмуртского университета. Серия Экономика и право. 2017. №2. С.50-59.
3. Пивкин, К. С. Прогнозирование ключевых показателей розничной сети во времени // Вестник Пермского университета. Серия "Экономика" = Perm University Herald. ECONOMY. 2017. Том 12. №4. С.592-608.
4. Пивкин, К. С. Реализация регрессионных методов прогнозирования товарного спроса с помощью языка R // Интеллектуальные системы в производстве. 2018. Том 16. №1. С. 15-25.

Публикации в остальных изданиях, материалы конференций

5. Пивкин, К. С. Использование математических методов прогнозирования в системе управления товарными запасами / К. С. Пивкин // Математические методы и интеллектуальные системы в экономике и образовании: сб. материалов всероссийской заочной науч.-практ. конф. – Ижевск, 2015. – С. 21-24.
6. Пивкин, К. С. Система управления товарными запасами на предприятии розничной торговли как объект экономико-математического исследования / К. С. Пивкин // Математические методы и интеллектуальные системы в экономике и образовании: сб. материалов всероссийской заочной науч.-практ. конф. – Ижевск, 2016. – С. 67-70.
7. Пивкин, К. С. Постановка задачи прогнозирования спроса как оценки математического ожидания / К. С. Пивкин // Математические методы и интеллектуальные системы в экономике и образовании: сб. материалов всероссийской заочной науч.-практ. конф. – Ижевск, 2017. – С. 23-26
8. Пивкин, К. С. Создание системы прогнозирования с помощью языка R на примере розничного предприятия / К. С. Пивкин // Перспективные информационные технологии (ПИТ 2017): труды Международной научно-технической конференции / под ред. С.А. Прохорова. – Самара: Издательство Самарского научного центра РАН, 2017. – С. 381-385.
9. Пивкин, К. С. Прогноз покупательского спроса как элемент системы управления товарными запасами / К. С. Пивкин // Материалы Международного молодежного научного форума «ЛОМОНОСОВ-2017» / Отв. ред. И.А. Алешковский, А.В. Андриянов, Е.А. Антипов. [Электронный ресурс] — М.: МАКС Пресс, 2017. – URL:

[https://lomonosov-
msu.ru/archive/Lomonosov_2017/data/section_13_10999.htm](https://lomonosov-msu.ru/archive/Lomonosov_2017/data/section_13_10999.htm)

10. Пивкин К.С. Кластеризация временных рядов как этап прогнозирования покупательского спроса / Математическое и компьютерное моделирование в экономике, страховании и управлении рисками: материалы VI Международной молодёжной научно-практической конференции. – Саратов: ООО Изд-во «Научная книга», 2017. – 264 с. –С. 165-169