

5. Skatkov A.V. et al. An adaptive method for detecting vulnerabilities in interfaces of unmanned vehicles in the infrastructure of a smart city. *Infokommunikacionnye tehnologii*, 2020, vol. 18, no. 1, pp. 45–50. (In Russ.)
6. Tkachenko K.S. Efficient digital support for changing requirements in manufacturing plants. *Infokommunikacionnye tehnologii*, 2020, vol. 18, no. 4, pp. 484–488. (In Russ.)
7. Tkachenko K.S., Skatkov I.A. Statistical modeling method for designing information-measuring control systems, taking into account the features of stochastic processes. *Sistemy kontrolja okruzhajuschej sredy*, 2020, no. 1 (39), pp. 46–53. (In Russ.)
8. Skatkov A.V., Tkachenko K.S. Statistical risk assessments in conditions of unauthorized perturbations of nodal traffic. *Sistemy kontrolja okruzhajuschej sredy*, 2016, no. 5 (25), pp. 41–46. (In Russ.)
9. Tkachenko K.S., Korepanova N.L. Detection of virus attacks in distributed environments and homogeneous networks of critical use. *Sistemy kontrolja okruzhajuschej sredy*, 2014, no. 20, pp. 98–101. (In Russ.)
10. Saichev A.I., Filimonov V.A., Tarakanova M.V. Estimate of the diffusion coefficient of a Wiener random process with uniform drift. *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo*, 2010, no. 5–1, pp. 61–66. (In Russ.)

Received 21.09.2021

НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 658.8; 65.011.56; 004

ИЗУЧЕНИЕ ВОЗМОЖНОСТЕЙ ПРЕДИКТИВНОЙ АНАЛИТИКИ ДАННЫХ FMCG

Павлюченко К.И.¹, Панфилов П.Б.¹, Горшков Г.С.²

¹ *Национальный исследовательский университет «Высшая школа экономики», Москва, РФ*

² *Московский финансово-юридический университет МФЮА, Москва, РФ*

E-mail: pavlyuchenkoki@gmail.com

В данной работе рассматривались возможности использования предиктивной аналитики данных в повышении эффективности бизнес-процессов на рынке товаров повседневного спроса FMCG, в частности для решения задач прогнозирования спроса на продукцию. Были проанализированы существующие бизнес-процессы в компаниях, FMCG и рассмотрены типовые инструменты для прогнозирования продаж, предлагаемые различными вендорами, включая такие, как SAP, RapidMiner, Azure ML Studio, SPSS. В качестве примера сбора и аналитической обработки данных, генерируемых бизнес-процессами на рынке FMCG, в работе были проанализированы данные по продажам ключевого клиента компании из табачной отрасли, созданы несколько регрессионных моделей прогнозирования продаж на основе использования аналитического продукта Azure ML Studio.

Ключевые слова: *предиктивный анализ данных, ритейл-бизнес, FMCG, прогнозирование продаж, регрессионная модель, Azure ML Studio*

Введение

Существуют различные рынки в зависимости от продаваемого продукта или услуги. Наиболее привычным для обычного потребителя является рынок «товаров повседневного спроса», известных также как «ширпотреб» (сокращение от «широкое потребление») или FMCG (по английской аббревиатуре термина Fast Moving Consumer Goods). Эта область рынка является высоко конкурентной, что подтверждается тем фактом, что существует множество транснациональных компаний, таких как Coca-Cola, Pepsi,

PMI, P&G, Unilever, Johnson&Johnson и другие. В прошлом году по данным NielsenIQ розничные продажи FMCG в России выросли на 3 % в денежном выражении, тогда как в 2019 году рост составил 3,9 % [1]. При этом потенциал роста за счет расширения торговых площадей и повышения цен оказался практически исчерпан. Этот факт, а также острая конкуренция заставляют крупнейших ритейлеров инвестировать в технологии, обеспечивающие преимущества, которых нет (пока) у других конкурентов [2]. Эти технологические решения могут быть основаны на самых разных принципах – психологических, маркетин-

говых, технических. Их главное назначение – это обеспечение более точного соответствия предложения продуктов потребностям покупателя. Например, технология компьютерного зрения уже сегодня позволяет сократить очереди на кассах в магазинах сети «Перекресток». Магазины компании «Лента» использует технологию обработки больших данных, чтобы знать, что вы, как покупатель, купите в следующий раз, когда пойдете в магазин. В компании «Магнит» технология машинного обучения позволяет всегда иметь в наличии именно те товары, которые нужны покупателям. Наиболее интересной технологией для анализа данных на рынке FMCG является предиктивная аналитика, поскольку она способна (потенциально) предсказывать неизвестные будущие события, отвечая на вопрос «Что может произойти?» на основе анализа накопленной информации о покупках и продажах товаров. Для этих целей используется множество методов: математическая статистика, моделирование, машинное обучение и другие направления «науки о данных» (Data Science), а также интеллектуальный анализ данных (Data Mining) [3]. Эти технологии позволяют предприятиям торговли не только накапливать данные, генерируемые бизнес-процессами, но и использовать их для улучшения бизнес-процессов.

Целью данной исследовательской работы является поиск путей улучшения эффективности бизнеса на рынке FMCG на основе применения методов и моделей предиктивной аналитики данных для решения задач прогнозирования продаж FMCG.

Для достижения указанной цели в работе решается ряд задач, включая:

- Анализ существующих бизнес-процессов и бизнес-моделей на рынке FMCG;
- Исследование методов предиктивной аналитики, которые используются для работы с данными, генерируемыми бизнес-процессами на рынке FMCG, и выбора методов и технических решений для прогнозирования продаж и планирования складских запасов продуктов для ретейла;
- Демонстрация примера использования предиктивной модели в процессе прогнозирования продаж на рынке FMCG на примере данных продаж продуктов определенной товарной группы FMCG.

Особенности бизнес-процессов рынка FMCG

Продукты FMCG характеризуются, как правило, коротким сроком службы, быстрым темпом потребления и многократными покупками со

стороны потребителей. В теории любой продукт с такими характеристиками может быть отнесен к рынку FMCG, включая продукты питания, напитки, косметику, моющие средства, батарейки, туалетную бумагу и другие. Главное характерное свойство и главный критерий здесь – это оборачиваемость или частота покупок, а это означает, что компания в данной отрасли должна работать с товарами очень быстро, иначе конкуренты сделают новый запуск продукта и перемянут к себе потребителя.

Иногда FMCG также называют PMCG (т. е. упакованные товары повседневного спроса). Как следует из названия, продукты упаковываются в отдельные небольшие единицы для продажи. Самые простые примеры продуктов этой категории включают упакованные продукты питания, средства личной гигиены, табак, алкоголь и напитки. Не случайно, что компании на рынке FMCG уделяют все больше внимания влиянию упаковки, брендинга и популяризации товаров этой категории.

Выделяют следующие важные атрибуты рынка товаров повседневного спроса FMCG:

- Короткий цикл оборота продукции;
- Доступ на рынок короткий и широкий;
- Рынок является ярким, т. е. в местах с высокой проходимостью и высоким качеством продукции устанавливаются наружные рекламные щиты для рекламы имиджа товара, а в магазинах проводятся демонстрации на месте, рекламные акции, распродажи со скидками и другие мероприятия;
- Удобство: потребители могут привычно совершать покупки поблизости;
- Визуализированные продукты: на потребителей легко влияет атмосфера магазина, когда они совершают покупку;
- Низкая лояльность к бренду: потребители легко меняют бренды среди аналогичных товаров.

Все эти ключевые моменты рынка обеспечивают то, что потребитель покупает продукты FMCG просто, быстро, импульсивно и эмоционально.

По предпочтительным бизнес-моделям и реализуемым бизнес-процессам компании FMCG в мире обычно относятся к розничной торговле онлайн или офлайн, поскольку это основные каналы продаж и связи с конечным потребителем для многих компаний-производителей. В таблицах 1 и 2 представлены Топ-10 брендов рынка FMCG в мире и в России [4; 5].

Исходя из данных в таблице 1, видно, что по охвату, присутствию на рынке и выбору потреби-

Таблица 1. Топ-10 брендов FMCG в мире

№	Компания	Бренд	CRPs* охват потребителей, млн чел.	Изменение, %	Доля на мировом рынке		Выбор потребителей	
					2019	2020	2019	2020
1	The Coca-Cola Company	Coca-Cola	6509	4	42,2	43	12,4	12,3
2	Colgate-Palmolive	Colgate	4311	-2	59,8	58,5	6,1	6
3	Unilever	Lifebuoy	3014	15	25,5	27,7	8,6	8,9
4	Nestlé	Maggi	3004	-4	32	33,9	8,1	7,2
5	PepsiCo Inc.	Lay's	2768	4	30,4	31,2	7,3	7,2
6	PepsiCo Inc.	Pepsi	2354	6	22,7	23,3	8,1	8,2
7	Indofood	Indomie	2221	0	6	6,1	30,9	29,5
8	Unilever	Dove	2033	1	37	37,1	4,5	4,5
9	Unilever	Sunsilk	1943	-4	23,5	23,4	7,2	6,8
10	Nestlé	Nescafé	1814	-3	22,7	23,2	6,9	6,4

Таблица 2. Топ-10 брендов FMCG в России

№	Компания	Бренд	CRPs* охват потребителей, млн чел.	Изменение, %	Доля на мировом рынке		Выбор потребителей	
					2019	2020	2019	2020
1	Danone	Простоквашино	487	-3	75,7	75,4	11,5	11,2
2	PepsiCo Inc.	Lay's	481	8	67,4	69,1	12,1	12,8
3	ОАО «Эфко продукты питания»	Слобода	304	-7	77	76	7,5	7,1
4	ЗАО Эссен Продакшн АГ	Махеев	286	10	72,6	77,2	6,3	6,5
5	The Coca-Cola Company	Coca-Cola	248	7	48,2	49,3	8,9	9,2
6	ОАО «Прогресс»	Фрутоняня	242	-4	43,5	43,7	9,5	9,1
7	Группа КДВ	Яшкино	236	-3	71	71,1	6,1	5,9
8	PepsiCo Inc.	Домик в деревне	236	-7	61,2	59,1	7,3	7,1
9	Mareven Food Holdings	Ролтон	232	5	64,1	65	6,1	6,3
10	Wm. Wrigley Jr. Company, Wrigley	Orbit	229	-10	52,7	49,2	8,9	8,5

телей лидируют транснациональные компании, в топ-3 список входят компании The Coca-Cola Company, Colgate-Palmolive, Unilever. Бренды этих компаний известны и присутствуют в большинстве крупных стран и относятся к пищевой или бытовой категории товаров.

Для России этот список будет немного другим, так как в каждой стране есть своя специфика. Как это видно по данным из таблицы 2, бренды по ключевым показателям в России относятся в основном к пищевой отрасли. Однако доля транснациональных компаний тут заметно ниже:

6 из 10 в России против 10 из 10 в мире. Это говорит о том, что потребитель в России по своему уникален и к нему нужен особый подход.

Но независимо от страны или региона все топ-10 продуктов потребитель с легкостью может найти в ближайшем супермаркете или на их сайте. Поэтому для FMCG-компаний важно распространять свою продукцию через рынок ретейла, т. е. рынок розничных продаж. Все продвижение и маркетинговые активности со стороны компаний-производителей в ретейле направлены на поддержание своего продукта в числе самых

продаваемых из категории. Это обусловлено тем, что если компании-производители продукта X и Y предлагают ретейлу одинаковую маржу с 1 продажи, то розничному бизнесу условно без разницы, какой продукт продавать. Таким образом, в сложных условиях, где потребитель может с легкостью выбрать новый бренд конкурента, бизнес должен понимать, сколько необходимо производить нового или существующего продукта на рынке. И так как производить больше, чем можно продать, неэффективно, то компании в FMCG-сфере должны как можно точнее прогнозировать спрос на свою продукцию.

Предиктивная аналитика в прогнозировании продаж товаров

В компаниях сферы FMCG существует множество бизнес-процессов, в которых можно использовать предиктивную аналитику, оптимизацию или автоматизацию. Например, логистика и цепочки поставок, управление запасами и бережливое производство, управлением кадрами. В каждом отдельном бизнес-процессе есть входные данные (начало процесса) и выходные данные (результат). Наиболее интересной для нашего исследования темой является прогнозирование продаж продукции, так как с одной стороны чрезмерное производство продукции порождает загрузку складов компании, что негативно сказывается на общем ее финансовом положении. С другой стороны, недопроизводство продукции можно оценить как упущенную прибыль из-за несвоевременных поставок в точки реализации продукта. Поэтому планирование и прогнозирование продаж/спроса – очень важный бизнес-процесс, который можно сделать еще более эффективным за счет внедрения моделей и методов машинного обучения и предиктивной аналитики.

Прогнозирование спроса является одной из проблем, которую можно решить за счет предиктивной аналитики. Согласно SAS Institute, предиктивная аналитика – это использование данных, статистических алгоритмов и методов машинного обучения для определения вероятности будущих результатов на основе исторических данных. Цель аналитики данных здесь состоит в том, чтобы не ограничиваться знанием статистики прошлого, а чтобы дать наилучшую оценку того, что произойдет в будущем. Предиктивная аналитика уже является одной из наиболее широко используемых технологий интеллектуальной автоматизации в мире. По данным Statista, более 80 % крупных предприятий внедряют предиктивную аналитику [6].

Предиктивная аналитика часто обсуждается в контексте больших данных, например инженерных данных, поступающих от датчиков, приборов и подключенных систем в бизнес. Бизнес-системы компании могут включать данные о транзакциях, результатах продаж, жалобах клиентов и маркетинговую информацию. Все чаще компании принимают решения на основе данных, основываясь на этой информации. Чтобы извлечь ценность из больших данных, предприятия применяют алгоритмы к большим массивам данных с помощью таких инструментов, как Hadoop и Spark [7]. Источники данных могут состоять из баз данных транзакций, журнальных файлов оборудования, изображений, видео, аудио, сенсорных и других типов данных. Инновации часто появляются благодаря объединению данных из нескольких источников. При наличии всех этих данных необходимы инструменты для извлечения информации и выявления тенденций. Методы машинного обучения используются для поиска закономерностей в данных и построения моделей, которые предсказывают будущие результаты. Существует множество алгоритмов машинного обучения, включая линейную и нелинейную регрессию, нейронные сети, деревья решений и другие алгоритмы.

В предыдущих работах авторы уже обсуждали и анализировали индустрию FMCG в своих странах. А также исследовали, как компании FMCG пытаются улучшить свои позиции на рынке. Одна из работ связана с рынком Китая и каналом электронной коммерции в отрасли [8]. Другая работа связана с рынком товаров массового спроса в Бангладеш и влиянием COVID-19 на HR-процессы [9]. Примеры работ показывают, что важно предварительно проанализировать, как компания может улучшить свой бизнес за счет внедрения новых цифровых возможностей. Например, компания может получить больше конкурентных преимуществ за счет внедрения IT-технологий, таких как предиктивная аналитика.

В последние годы предиктивной аналитике уделяется большое внимание в связи с развитием вспомогательных технологий, особенно в области больших данных и машинного обучения. Предиктивная аналитика помогает командам в таких различных отраслях, как финансы, здравоохранение, фармацевтика, автомобилестроение, аэрокосмическая промышленность и производство [10]. Машинное обучение – это категория алгоритмов, которая позволяет программным приложениям становиться более точными в прогнозировании результатов без явного програм-

мирования. Основная предпосылка машинного обучения заключается в построении моделей и применении алгоритмов, которые могут получать входные данные и использовать статистический анализ для прогнозирования выходных данных, обновляя их по мере поступления новых данных. Эти модели можно применять в различных областях и обучать их в соответствии с ожиданиями руководства, чтобы предпринимать точные шаги для достижения целей организации [11]. Предиктивная аналитика, включая прогнозирование, является важным инструментом для функции продаж [12]. Прогнозы дают представление об анализе клиентов и счетов [13], коэффициентах конверсии, достижении квоты, стратегии командных продаж [14], подборе и оценке персонала [15].

Предиктивная аналитика – это категория анализа данных, предназначенная для составления прогнозов на основе исторических данных с целью моделирования будущих сценариев с использованием аналитических методов, таких как статистическое моделирование и машинное обучение или глубокое обучение. Используя предиктивную аналитику, организация может обнаружить тенденции и адаптировать свою политику к тому, что может произойти [16].

Для обзора всех возможностей предиктивной аналитики мы рассмотрим проект с открытым исходным кодом и код на языке Python. Язык Python стал самым популярным языком машинного обучения благодаря своей простоте, читабельности и расширяемости. Расширяя библиотеку NumPy, можно добиться быстрой обработки массивов. В то же время Python может напрямую реализовать алгоритм машинного обучения, расширив фреймворк TensorFlow [17].

Методы решения задач предиктивной аналитики данных FMCG

Подходы и методы, используемые для проведения прогнозных анализов, можно разделить на два вида: методы классификации и методы кластеризации.

Классификация объекта – номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту [18]. Обучение классификатора – процесс построения алгоритма в случае, когда задано конечное множество объектов, для которых известно, к каким классам они относятся. Это множество называется выборкой. Классовая принадлежность остальных объектов неизвестна. Это также называется «обучением с учителем».

Кластеризация – процесс разбиения заданной выборки объектов (наблюдений) на непересекающиеся подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались [18]. Одной из целей кластеризации является понимание данных путем выявления кластерной структуры. Разбиение наблюдений на группы схожих объектов позволяет упростить дальнейшую обработку данных и принятия решений, применяя к каждому кластеру свой метод анализа. Данный метод машинного обучения относится к «обучению без учителя», так как у данных нет заранее возможных классов.

Рассмотрим более подробно последовательность действий при внедрении модели предиктивной аналитики в бизнес-процесс.

- Определить бизнес-цели моделирования;
- Выбор/получение данных;
- Подготовить данные;
- Анализ и преобразование переменных;
- Выбор и разработка моделей с учетом потребностей бизнеса;
- Тренировка модели (на тренировочной выборке данных);
- Валидация модели (на тестовой выборке);
- Оптимизация модели (если требуется);
- Внедрение и масштабирование в бизнес-процессе.

Эти шаги можно сгруппировать в три основные категории: подготовка данных, прогнозирование модели и развертывание [16]. Подготовка данных – это самостоятельная деятельность, которая преобразует разрозненные, необработанные и беспорядочные данные в четкое, последовательное представление. Этот процесс включает в себя исследование, очистку, преобразование, организацию и сбор данных. Подготовка данных важна, но требует времени; команды по работе с данными тратят до 80 % своего времени на преобразование необработанных данных в высококачественные результаты, готовые к анализу. Данные могут поступать из разных источников и объединяться в одном файле. На этапе построения модели прогнозирования мы строим модель, которая будет учиться на исторических и текущих данных, чтобы предсказывать будущее состояние системы. Как уже говорилось выше, любой вид алгоритма, который может быть использован, зависит от типа данных и бизнес-цели. Созданная прогностическая модель должна быть протестирована и подтверждена. После завершения всех шагов, связанных с настройкой и отладкой модели, она должна быть запущена в организации и

Таблица 3. Инструменты для анализа данных и машинного обучения

Продукт	Применение
SAP Analytics Cloud	Наиболее сбалансированный программный продукт с возможностью постройки сложных предиктивных моделей
SAS Advanced Analytics	Глобальный продукт для крупных компаний, подходит для любых задач в области анализа данных
RapidMiner	Бесплатный сервис с автоподбором лучшей предиктивной модели на основе используемых данных
IBM SPSS	Лучше всего подходит для работы с данными при исследованиях
Azure ML Studio	Имеет широкий выбор моделей, подходит для специалистов без знания программирования

быть доступной для всех лиц, принимающих решения, или людей, вовлеченных в процесс.

Для анализа данных и постройки моделей машинного обучения в бизнесе используются различные инструменты и готовые пакетные решения. Наиболее известные из них представлены в таблице 3, где даны краткие характеристики каждого инструмента или пакетного решения, которое компания FMCG может внедрить и использовать в своих бизнес-процессах.

SAP Analytics Cloud – это облачное решение, объединяющее функции планирования, бизнес-аналитики и прогнозного анализа. Оно упрощает финансовое планирование и анализ и позволяет сотрудникам исследовать данные и совместно работать в общем контексте в реальном времени [19].

SAS Visual Analytics – это высокопроизводительное решение in-memory для анализа больших объемов данных. Оно позволяет пользователям обнаруживать закономерности, определять направления дальнейшего анализа, передавать полученные визуальные результаты в виде отчетов для web- или мобильных устройств [20].

RapidMiner – платформа, которая работает по принципу клиент-серверной модели, причем сервер может быть размещен как в локальной, так и в облачной инфраструктуре. Основная особенность этой платформы – отсутствие необходимости в написании рабочих кодов, что существенно повышает скорость обработки данных и уменьшает количество ошибок [21].

Alteryx разработал среду для перетаскивания и смешивания данных и расширенной аналитики, которая помогает аналитику получить необходимую информацию в течение нескольких часов, а не недель. Программа делает это с помощью

широкого спектра инструментов, которые дают доступ, готовят, анализируют и выводят данные быстрее и проще. Каждая вкладка представляет собой образец определенных инструментов Designer Alteryx, которые поддерживают полный спектр возможностей в пределах Alteryx [22].

Платформа SPSS компании IBM предлагает передовые инструменты статистического анализа, обширную библиотеку алгоритмов машинного обучения, анализа текста, расширения компонентов с открытым кодом, интеграции с большими данными и беспрепятственного внедрения в приложения. Благодаря простоте эксплуатации, гибкости и масштабируемости SPSS отлично подходит пользователям с любым уровнем подготовки. Более того, SPSS подходит для реализации проектов любого объема и сложности, направленных на поиск новых возможностей, повышение эффективности и снижение рисков [23].

Платформа Azure ML от компании Microsoft представлена двумя компонентами: Azure ML Studio – средой разработки, доступной через браузер и предоставляющей визуальный интерфейс для создания моделей в стиле Drag&Drop, а также web-сервисами Azure ML для использования моделей в промышленном решении. Ядро Azure ML основано на самых современных реализациях алгоритмов машинного обучения, которые используются внутри Microsoft, а также предоставляет сотни пакетов для таких популярных языков, как R и Python [24].

Пример предиктивной аналитики данных FMCG для прогнозирования продаж

Для дальнейшего примера мы будем использовать одно из описанных выше решений, а именно: Azure ML Studio. В качестве тестовых данных для построения и настройки предиктивной модели будут использоваться количественные данные продаж клиента из России, осуществляющего реализацию продукции табачной компании X рынка FMCG. Особенность этого сегмента рынка FMCG обусловлена тем, что компания-производитель не может осуществлять такую же маркетинговую активность, как в случае других товаров FMCG [25]. Тем не менее прогнозирование спроса данной продукции является приоритетной задачей для производителя. Также персональные данные о покупателях получить можно только через третьих лиц, что не всегда качественно отражается на прогнозах компании. Поэтому для работы модели можно использовать данные из открытых источников.

Таблица 4. Тестовый набор данных продаж ключевого клиента

Год	Месяц	Ключевая ставка	Обменный курс руб./доллар	Уровень инфляции	Отгрузки в 10 млн штук
2013	1	8,25	30,2414	7,07	31,41172
2013	2	8,25	30,1245	7,28	31,90266
2013	3	8,25	30,7769	7,02	38,06697
2013	4	8,25	31,3169	7,23	39,83486
2013	5	8,25	31,3285	7,38	42,89173
2013	6	8,25	32,2822	6,88	40,53027
2013	7	8,25	32,64	6,45	41,44613
2013	8	8,25	33,0004	6,49	41,40899
2013	9	5,5	32,5091	6,13	39,06895
2013	10	5,5	32,125	6,25	38,64342
2013	11	5,5	32,6874	6,48	36,4045
2013	12	5,5	32,8658	6,45	39,88357
2014	1	5,5	33,6429	6,05	36,44582
2014	2	5,5	35,2366	6,2	32,81735
2014	3	7	36,2344	6,92	32,88184
2014	4	7	35,6656	7,33	32,77991
2014	5	7,5	34,7221	7,59	35,33578
2014	6	7,5	34,3936	7,8	35,11416
2014	7	8	34,4258	7,45	39,64588
2014	8	8	36,1098	7,56	40,76325
2014	9	8	37,9861	8,03	38,15826
2014	10	8	40,7457	8,3	38,0874
2014	11	9,5	46,3379	9,07	36,26518
2014	12	17	54,4367	11,36	40,37127
2015	1	17	65,2869	14,97	35,17325
2015	2	15	64,2972	16,71	34,64783
2015	3	14	60,6649	16,93	39,41995
2015	4	14	52,363	16,42	39,6544
2015	5	12,5	50,3419	15,78	42,6393
2015	6	11,5	54,3683	15,29	42,8885
2015	7	11,5	56,9774	15,64	43,4949
2015	8	11	65,0169	15,77	42,86166
2015	9	11	66,5954	15,68	40,44601
2015	10	11	62,7061	15,59	41,89715
2015	11	11	64,912	14,98	40,94255
2015	12	11	70,2244	12,91	45,25844
2016	1	11	76,5845	9,77	40,69214
2016	2	11	77,1326	8,06	41,27969
2016	3	11	70,2305	7,26	45,43016
2016	4	11	66,4756	7,24	45,7325
2016	5	11	65,9681	7,3	47,78067
2016	6	10,5	65,1339	7,48	47,06741
2016	7	10,5	64,1127	7,21	49,43757
2016	8	10,5	64,8139	6,84	48,39184
2016	9	10	64,7579	6,42	43,63153

Продолжение таблицы 4

2016	10	10	62,4583	6,09	43,96002
2016	11	10	64,1833	5,76	42,27497
2016	12	10	61,6368	5,38	47,19977
2017	1	10	59,6526	5,02	43,4317
2017	2	10	58,0967	4,59	42,72794
2017	3	9,75	58,2437	4,25	50,31987
2017	4	9,25	56,3131	4,13	52,01197
2017	5	9,25	56,756	4,09	56,61702
2017	6	9,25	57,4437	4,35	57,19999
2017	7	9	59,5787	3,86	60,60408
2017	8	9	59,799	3,29	60,97246
2017	9	9	57,7192	2,96	57,6521
2017	10	8,5	57,6869	2,73	57,11693
2017	11	8,25	59,0061	2,5	54,07691
2017	12	8,25	58,6932	2,52	58,42475
2018	1	7,75	56,5925	2,21	52,68072
2018	2	7,5	56,6278	2,2	48,74953
2018	3	7,5	57,0113	2,36	53,69429
2018	4	7,25	61,5539	2,41	54,69175
2018	5	7,25	62,3033	2,42	59,86509
2018	6	7,25	62,7565	2,3	59,1236
2018	7	7,25	62,9471	2,5	66,30797
2018	8	7,25	66,8932	3,07	61,99188
2018	9	7,5	68,0447	3,39	57,48137
2018	10	7,5	65,7492	3,55	58,53678
2018	11	7,5	66,0499	3,83	55,71769
2018	12	7,75	66,7848	4,27	58,44602

Используя последовательность действий, описанных выше, следует начать с определения бизнес-целей. Бизнес-цель компании X – это улучшить точность прогнозирования продаж, чтобы точнее определять спрос на свою продукцию в условиях рыночной неопределенности.

Далее необходимо определить, какие данные для создания предиктивной модели необходимо использовать. Были выбраны данные по количественным продажам продукции компании X за период 2013–2018 гг. Также можно предположить, что существует зависимость между макроэкономическими показателями и продажами табачной продукции ключевого клиента. Для данного примера были использованы такие данные, как ключевая ставка, курс рубля к доллару, инфляция (таблица 4). Этот набор данных был назван dataset и сохранен в формате csv. Так как никаких дополнительных преобразований с данными происходить не будет, то далее мы переходим к выбору и разработке предиктивной модели.

В качестве основного программного инструмента по анализу данных и созданию моделей было выбрано готовое решение Microsoft Azure Machine Learning Studio [26], или, сокращенно, Azure ML Studio. Эта достаточно мощная платформа аналитики больших данных позволяет реализовать настройку, обучение и тестирование моделей машинного обучения без необходимости написания самого программного кода модели, а с помощью элементов графического интерфейса пользователя, как это представлено на рисунке 1.

Так как прогнозирование продаж есть проблема правильной классификации объекта в числовом виде, то выбор оптимальной модели проводился среди регрессионных моделей машинного обучения. В данной работе рассматривались следующие пять основных моделей.

Байесовская линейная регрессия – это подход в линейной регрессии, в котором статистический анализ проводится в контексте байесовского вывода. Когда регрессионная модель характери-

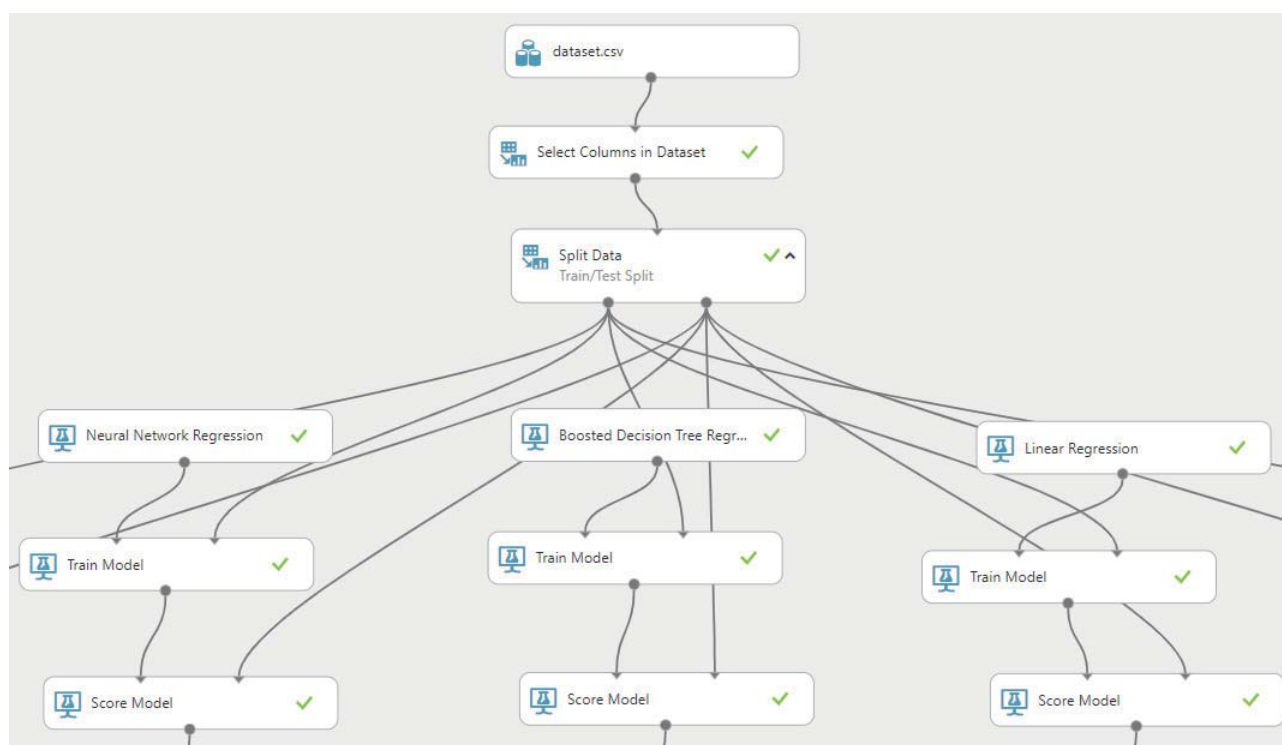


Рисунок 1. Часть интерфейса из платформы Azure ML Studio

зается ошибками, имеющими нормальное распределение, и принимается определенная форма априорного распределения, доступны явные результаты для апостериорных распределений вероятностей параметров модели [27]. Задаваемые параметры модели в Azure ML Studio в этом случае следующие:

- L1 regularization weight = 1;
- Allow unknown categorical levels = True.

Нейросетевая регрессия – модель является контролируемым методом обучения, основанном на принципах построения искусственных нейронных сетей с несколькими слоями узлов обработки данных [28]. Задаваемые параметры модели здесь следующие:

- Hidden layer specification = «fully-connected case»;
- Number of hidden nodes = 300;
- Learning rate = 0,01, 0,02, 0,04;
- Number of iterations = 20, 40, 80, 160, 320;
- The initial learning weights diameter = 0,1;
- The momentum = 0;
- The type of normalizer = «Min-Max normalizer».

Модуль регрессии повышающегося дерева принятия решений – эта модель используется для создания ансамблей деревьев регрессии путем повышения. Повышение означает, что каждое дерево зависит от предыдущих деревьев. Алгоритм обучается путем подгонки остатка предыдущего дерева. Таким образом, «бустинг»

в наборе деревьев принятия решений обычно обеспечивает повышение точности с небольшим риском снижения покрытия [29]. Параметры модели в этом случае следующие:

- Maximum number of leaves per tree = 40;
- Minimum number of samples per leaf node = 10;
- Learning rate = 0,1;
- Total number of trees constructed = 100;
- Random number seed = 1;
- Allow unknown categorical levels = True.

Линейная регрессия – это общий статистический метод, который был реализован в машинном обучении и дополнен многими новыми методами для подгонки строки и измерения ошибок. Простыми словами, регрессия связана с прогнозированием числовых целевых значений [30]. Параметры модели в этом случае следующие:

- Solution method = «Online Gradient Descent»;
- Learning rate = 0,025; 0,05; 0,1; 0,2;
- Number of training epochs = 1, 10, 100;
- L2 regularization weight = 0,001; 0,01; 0,1;
- Normalize features, Average final hypothesis, Decrease learning rate = True;
- Random number seed = 1;
- Allow unknown categorical levels = True.

Деревья принятия решений – это непараметрические модели, выполняющие последовательность простых тестов для каждого экземпляра данных при обходе древовидной структуры двоичных данных до достижения конечного узла

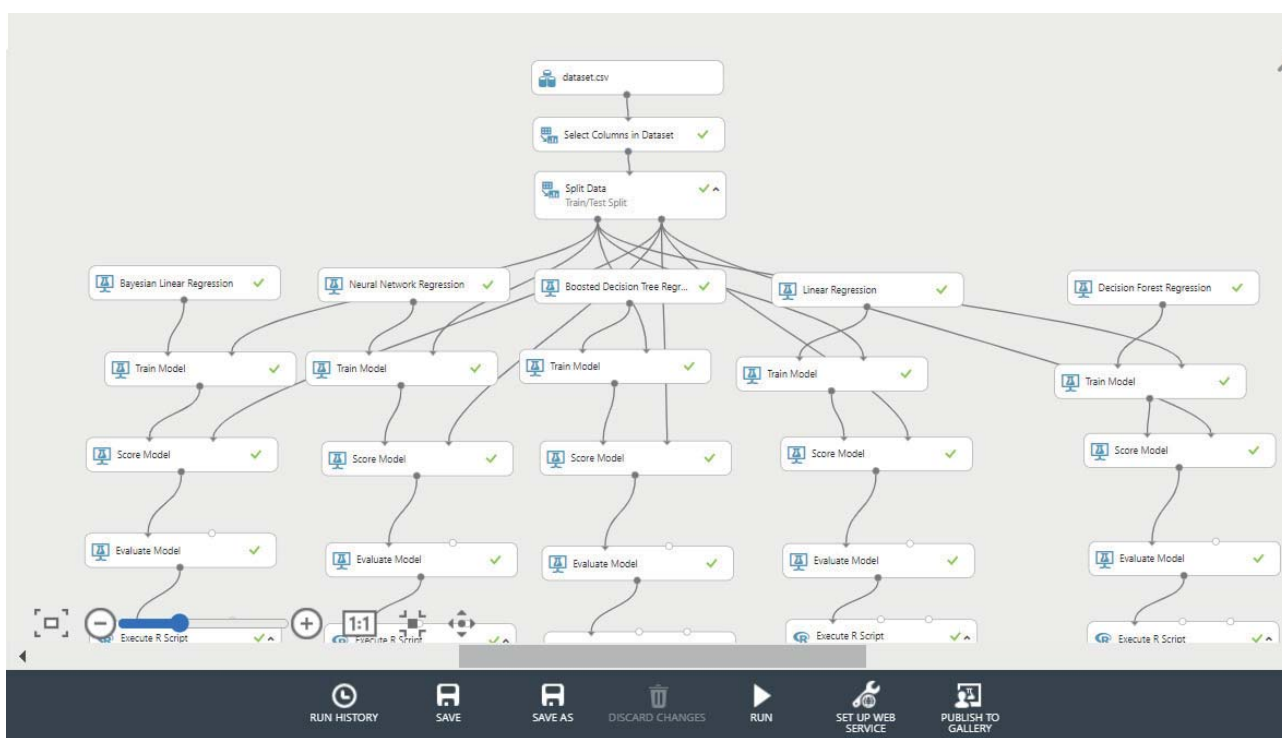


Рисунок 2. Архитектура построенных моделей в Azure ML Studio, ч.1

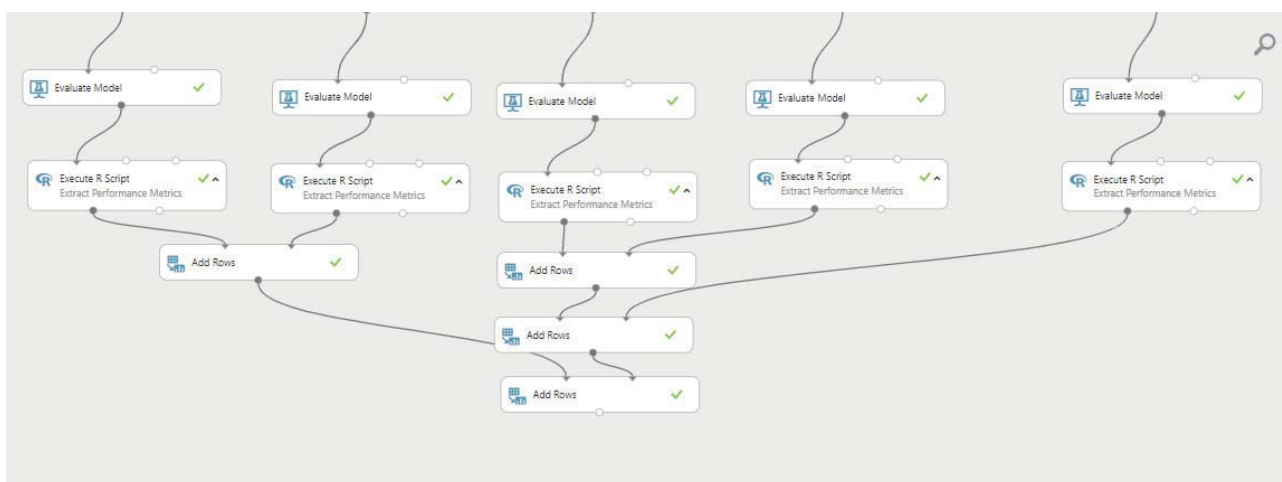


Рисунок 3. Архитектура построенных моделей в Azure ML Studio, ч.2

(решения) [31]. Параметры модели в этом случае следующие:

- Resampling method = «Bagging»;
- Number of decision trees = 1, 8, 32;
- Maximum depth of the decision trees = 1, 16, 64;
- Number of random splits per node = 1, 128, 1024;
- Minimum number of samples per leaf node = 1, 4, 16;
- Allow unknown categorical levels = True.

Последовательность действий при работе с инструментарием платформы Azure следующая.

1. Загрузить данные на платформу.
2. Выбрать компонент Split data и выбрать те колонки таблицы набора данных, которые не

представляют ценности для анализа (в нашем тестовом примере была исключена колонка с данными показателя года).

3. Выбрать разделение набора данных на тестовую и тренировочную выборку данных. В нашей модели тренировочная выборка составляет 20 % от набора данных, и 80 % используются для тренировки модели.

4. Оценка модели: показатели моделей можно оценить по среднеквадратическому отклонению.

5. Все показатели пяти сравниваемых моделей собираются скриптом на языке R для обзора и выбора наиболее точной модели прогнозирования.

Таблица 5. Результаты разработанных и протестированных моделей прогнозирования продаж

	Средний модуль отклонения	Средний квадрат отклонения	Относительная абсолютная погрешность	Средний квадрат относительной абсолютной погрешности	Коэффициент детерминации
Байесовская линейная регрессия	4,40	5,29	0,59	0,41	0,59
Регрессия нейронной сети	3,28	3,76	0,44	0,21	0,79
Модуль регрессии повышающегося дерева принятия решений	2,24	2,61	0,30	0,10	0,90
Линейная регрессия	19,75	21,26	2,63	6,64	-5,64
Деревья принятия решений	2,10	2,47	0,28	0,09	0,91

rows 14
columns 7

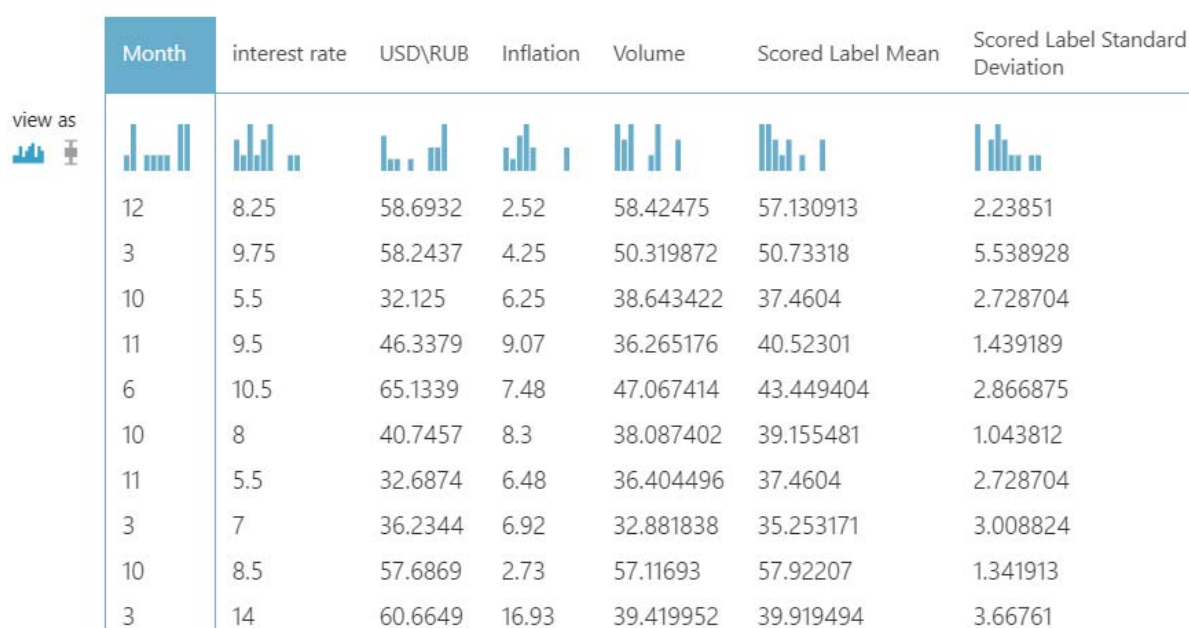


Рисунок 4. Пример результата работы предиктивной модели

6. Сохранение результатов моделирования, оценка полученных данных и финальный выбор модели.

Результаты работы пяти регрессионных моделей на тестовом наборе данных сведены в таблице 5.

Из таблицы 5 видно, что наиболее успешно прогнозирует модель на основе деревьев принятия решений. Имея наименьшую ошибку и максимальный коэффициент детерминации, данная модель может быть использована для реального прогнозирования будущих значений на основе выбранных признаков.

Результаты моделирования, представленные на рисунке 4, показывают, что в некоторых моментах модель довольно точно описывает и прогнозирует спрос на продукцию, но случаются и

сильные отклонения, большие чем средняя ошибка. Это может быть связано с тем, что модель оказалась сильно завязана на макроэкономические события, которые напрямую на клиента не повлияли, но модель их тем не менее учитывала в процессе выработки предсказаний.

Выводы

Как было отмечено в начале работы, российский рынок FMCG имеет свои характеристики и особенности, но и в этом случае с помощью учета макроэкономических факторов может быть выполнен последовательный процесс по разработке, анализу и оценке предиктивной модели для предсказания ситуации на рынке, прежде всего в отношении продаж специфических продуктов. Нами было продемонстрировано, как

предсказательная аналитика может работать на данных продаж табачной продукции как типичного примера пакетированных товаров FMCG. В результате была построена модель, способная прогнозировать количество продаваемой продукции табачной компании X через ключевого клиента со средней ошибкой 2,1 %, что составляет величину, меньшую 4,3 % от среднего значения тестовой выборки. Коэффициент детерминации 0,91 также высок, что является хорошим результатом. Также показатели других регрессионных моделей обнаруживают, что для более точной оценки и прогнозирования продаж у клиента необходимо рассматривать дополнительные источники информации и данных. Этот результат может быть внедрен непосредственно в бизнес-процесс планирования и прогнозирования продаж на будущие периоды. С помощью инструмента аналитики данных Azure ML Studio можно и дальше улучшать показатели выбранной предиктивной регрессионной модели за счет обогащения ее данными о клиенте. В ходе дальнейших практических исследований планируются внедрение данной модели в бизнес-процесс компании FMCG и оценка его эффективности на основе, в частности, анализа показателей загруженности складских помещений и запасов продукта в абсолютных числах.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-07-00958.

Литература

1. В 2020 году рынок FMCG в России замедлил рост до 3 %. URL: <http://www.finmarket.ru/news/5407578> (дата обращения: 01.11.2021).
2. 10 успешных кейсов внедрения технологий в ритейле. URL: <https://rb.ru/longread/retail-new-tech/> (дата обращения: 01.11.2021).
3. Вичугова А. Какая бывает аналитика: предиктивная, описательная и еще 2 вида аналитики больших данных. URL: <https://www.bigdataschool.ru/blog/types-of-data-analytics.html> (дата обращения: 01.11.2021).
4. Explore the data. Access the data and rankings from all the countries and sectors. URL: <https://www.kantar.com/campaigns/brand-footprint/explore-the-data> (дата обращения: 01.11.2021).
5. Фролов Д. Топ-50 FMCG-брендов в России. Исследование Nielsen. URL: <https://www.sostav.ru/publication/top-50-fmcg-brendov-v-rossii-issledovanie-nielsen-41081.html> (дата обращения: 01.11.2021).
6. Алексей PostMonitor. Предиктивная аналитика в маркетинге: где применяется, какой эффект можно получить. URL: <https://vc.ru/marketing/156155-prediktivnaya-analitika-v-marketinge-gde-primenyaetsya-kakoy-effekt-mozhno-poluchit> (дата обращения: 01.11.2021).
7. Predictive Analytics. 3 Things You Need to Know. URL: <https://www.mathworks.com/discovery/predictive-analytics.html> (дата обращения: 01.11.2021).
8. Shiqian Yu. Economic Analysis of the FMCG Industry in China (Fast Moving Consumer Goods). URL: <https://webthesis.biblio.polito.it/17759/1/tesi.pdf> (дата обращения: 01.11.2021).
9. Tania Akter. Impact of COVID-19 on Human Resource Management Practices of FMCG Industry in Bangladesh. URL: <http://dspace.uui.ac.bd/handle/52243/2029> (дата обращения: 01.11.2021).
10. Kumar V., Garg M.L. Predictive analytics: A review of trends and techniques // International Journal of Computer Applications. 2018. Vol. 182, no. 1. P. 31–37. DOI: <https://doi.org/10.5120/ijca2018917434>
11. Malik N., Singh K. Sales Prediction Model for Big Mart. URL: https://www.researchgate.net/publication/344099746_SALES_PREDICTION_MODEL_FOR_BIG_MART (дата обращения: 01.11.2021).
12. An information system for sales team assignments utilizing predictive and prescriptive analytics / J.K.V. Bischhoffshausen [et al.] // 2015 IEEE 17th Conference on Business Informatics. 2015. Vol. 1. P. 68–76. DOI: <https://doi.org/10.1109/CBI.2015.38>
13. Gilliland M., Tashman L., Sglavo U. Business forecasting: Practical problems and solutions // International Journal of Forecasting. 2017. DOI: <https://doi.org/10.1016/j.ijforecast.2017.06.002>
14. Prescriptive analytics for allocating sales teams to opportunities / B. Kawas [et al.] // 2013 IEEE 13th International Conference on Data Mining Workshops. 2013. P. 211–218. DOI: <https://doi.org/10.1109/ICDMW.2013.156>
15. Green D. Episode 16: McKinsey's Approach to Data-Driven HR (Interview with Keith McNulty, Global Director of People Analytics and Measurement at McKinsey). URL: <https://www.myhrfuture.com/digital-hr-leaders-podcast/2019/10/1/mckinseys-approach-to-data-driven-hr> (дата обращения: 01.11.2021).

16. Henrys K. Role of Predictive Analytics in Business. 2021. 13 p. URL: <https://ssrn.com/abstract=3829621> (дата обращения: 01.11.2021).
17. Shen G., Liu Q. Performance Analysis of Linear Regression Based on Python // Communications in Computer and Information Science. 2020. Vol 1227. DOI: https://doi.org/10.1007/978-981-15-6113-9_80
18. Черезов Д.С., Тюкачев Н.А. Обзор основных методов классификации и кластеризации данных // Вестник ВГУ, Серия: Системный анализ и информационные технологии. 2009. № 2. С. 25–29.
19. SAP BusinessObjects Cloud. URL: <https://www.id-mt.ru/produkty/sap/sap-businessobjects-cloud/> (дата обращения: 01.11.2021).
20. SAS Visual Analytics // Tadviser. URL: https://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:SAS_Visual_Analytics (дата обращения: 01.11.2021).
21. Платформа RAPIDMINER для анализа данных // Центр развития компетенций в области бизнес-информатики Высшей школы бизнеса НИУ ВШЭ. URL: <https://hsbi.hse.ru/articles/platforma-rapidminer-dlya-analiza-dannykh/> (дата обращения: 01.11.2021).
22. Инструменты Alteryx Designer. URL: <https://biconsult.ru/products/instrumenty-alteryx-designer> (дата обращения: 01.11.2021).
23. Приложения IBM SPSS. URL: <https://www.ibm.com/ru-ru/analytics/spss-statistics-software> (дата обращения: 01.11.2021).
24. Microsoft Azure // Tadviser. URL: https://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:Microsoft_Azure (дата обращения: 01.11.2021).
25. С 28.01.2021 года вступят в силу изменения в Федеральный закон «О рекламе». URL: <http://kurgan.fas.gov.ru/news/14080> (дата обращения: 01.11.2021).
26. Создавайте, обучайте и развертывайте модели машинного обучения с бесплатной учетной записью Azure. URL: <https://azure.microsoft.com/ru-ru/free/machine-learning/> (дата обращения: 01.11.2021).
27. Minka T.P. Bayesian Linear Regression. URL: <https://tminka.github.io/papers/minka-linear.pdf> (дата обращения: 01.11.2021).
28. Регрессия нейронной сети // Azure. Машинное обучение. URL: <https://docs.microsoft.com/ru-ru/azure/machine-learning/algorithm-module-reference/neural-network-regression> (дата обращения: 01.11.2021).
29. Модуль регрессии повышающегося дерева принятия решений // Azure. Машинное обучение. URL: <https://docs.microsoft.com/ru-ru/azure/machine-learning/algorithm-module-reference/boosted-decision-tree-regression> (дата обращения: 01.11.2021).
30. Модуль линейной регрессии // Azure. Машинное обучение. URL: <https://docs.microsoft.com/ru-ru/azure/machine-learning/algorithm-module-reference/linear-regression> (дата обращения: 01.11.2021).
31. Модуль регрессии леса принятия решений // Azure. Машинное обучение. URL: <https://docs.microsoft.com/ru-ru/azure/machine-learning/algorithm-module-reference/decision-forest-regression> (дата обращения: 01.11.2021).

Получено 08.11.2021

Павлюченко Константин Игоревич, магистрант департамента бизнес-информатики Высшей школы бизнеса Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ). 119049, Российская Федерация, г. Москва, ул. Шаболовка, д. 28/11, стр. 4. Тел. +7 968 043-03-33. E-mail: pavlyuchenkoki@gmail.com

Панфилов Петр Борисович, к.т.н., доцент, профессор департамента бизнес-информатики Высшей школы бизнеса НИУ ВШЭ. 119049, Российская Федерация, г. Москва, ул. Шаболовка, д. 28/11, стр. 4. Тел. +7 915 403-77-08. E-mail: ppanfilov@hse.ru

Горшков Георгий Сергеевич, к.т.н., доцент, проректор по информационным технологиям Московского финансово-юридического университета (МФЮА). 117342, Российская Федерация, г. Москва, ул. Введенского, 1А. Тел. +7 985 762-37-32. E-mail: Gorshkov.E@mfua.ru

EXPLORING THE CAPABILITIES OF PREDICTIVE DATA ANALYTICS IN FMCG INDUSTRY

Pavlyuchenko K.I.¹, Panfilov P.B.¹, Gorshkov G.S.²

¹ National Research University Higher School of Economics, Moscow, Russian Federation

² Moscow University of Finance and Law MFUA, Moscow, Russian Federation

E-mail: pavlyuchenkoki@gmail.com

This research paper considered the possibilities of using predictive analytics in the FMCG market to increase business process efficiency, in particular in the field of demand forecasting. The existing business processes in FMCG companies were considered as well as existing tools for sales forecasting such as SAP, RapidMiner, Azure ML Studio, SPSS and others. As an illustrative example of FMCG business process data collection and analytics, the data on the sales of the key customer of the company from the tobacco industry was analyzed and several regression models of sales forecasting were created on the basis of the use of the Azure ML Studio toolset.

Keywords: predictive analytics, retail, FMCG, sales forecasting, regression model, Azure ML Studio

DOI: 10.18469/ikt.2021.19.4.08

Pavyuchenko Konstantin Igorevich, Graduate School of Business of the National Research University Higher School of Economics, 28/11, building 4, Shabolovka Street, Moscow, 119049, Russian Federation; Graduate student of Business Informatics Department. Tel. +7 968 043-03-33. E-mail: pavlyuchenkoki@gmail.com

Panfilov Peter Borisovich, Graduate School of Business of the National Research University Higher School of Economics, 28/11, building 4, Shabolovka Street, Moscow, 119049, Russian Federation; Professor of Business Informatics Department, PhD of Technical Sciences, Associate Professor. Tel. +7 915 403-77-08. E-mail: ppanfilov@hse.ru

Gorshkov Georgy Sergeevich, Moscow University of Finance and Law MFUA, 1A, Vvedensky Street, Moscow, 117342, Russian Federation; Vice-Rector for Information Technology, PhD of Technical Sciences, Associate Professor. Tel. +7 985 762-37-32. E-mail: Gorshkov.E@mfua.ru

References

1. In 2020, the FMCG market in Russia slowed down growth to 3 %. URL: <http://www.finmarket.ru/news/5407578> (accessed: 01.11.2021). (In Russ.)
2. 10 successful cases of technology implementation in retail. URL: <https://rb.ru/longread/retail-new-tech/> (accessed: 01.11.2021). (In Russ.)
3. Vichugova A. What is analytics like: predictive, descriptive and 2 more types of big data analytics. URL: <https://www.bigdataschool.ru/blog/types-of-data-analytics.html> (accessed: 01.11.2021). (In Russ.)
4. Explore the data. Access the data and rankings from all the countries and sectors. URL: <https://www.kantar.com/campaigns/brand-footprint/explore-the-data> (accessed: 01.11.2021).
5. Frolov D. Top 50 FMCG brands in Russia. Nielsen research. URL: <https://www.sostav.ru/publication/top-50-fmcg-brendov-v-rossii-issledovanie-nielsen-41081.html> (accessed: 01.11.2021). (In Russ.)
6. Alexey PostMonitor. Predictive analytics in marketing: where it is applied, what effect can be obtained. URL: <https://vc.ru/marketing/156155-prediktivnaya-analitika-v-marketinge-gde-primenyaetsya-kakoy-effekt-mozhno-poluchit> (accessed: 01.11.2021). (In Russ.)
7. Predictive Analytics. 3 Things You Need to Know. URL: <https://www.mathworks.com/discovery/predictive-analytics.html> (accessed: 01.11.2021).
8. Shiqian Yu. Economic Analysis of the FMCG Industry in China (Fast Moving Consumer Goods). URL: <https://webthesis.biblio.polito.it/17759/1/tesi.pdf> (accessed: 01.11.2021).
9. Tania Akter. Impact of COVID-19 on Human Resource Management Practices of FMCG Industry in Bangladesh. URL: <http://dspace.uui.ac.bd/handle/52243/2029> (accessed: 01.11.2021).

10. Kumar V., Garg M.L. Predictive analytics: A review of trends and techniques. *International Journal of Computer Applications*, 2018, vol. 182, no. 1, pp. 31–37. DOI: <https://doi.org/10.5120/ijca2018917434>
11. Malik N., Singh K. Sales Prediction Model for Big Mart. URL: https://www.researchgate.net/publication/344099746_SALES_PREDICTION_MODEL_FOR_BIG_MART (accessed: 01.11.2021).
12. Bischhoffshausen J.K.V. et al. An information system for sales team assignments utilizing predictive and prescriptive analytics. *2015 IEEE 17th Conference on Business Informatics*, 2015, vol. 1, pp. 68–76. DOI: <https://doi.org/10.1109/CBI.2015.38>
13. Gilliland M., Tashman L., Sglavo U. Business forecasting: Practical problems and solutions. *International Journal of Forecasting*, 2017. DOI: <https://doi.org/10.1016/j.ijforecast.2017.06.002>
14. Kawas B. et al. Prescriptive Analytics for Allocating Sales Teams to Opportunities. *2013 IEEE 13th International Conference on Data Mining Workshops*, 2013, pp. 211–218. DOI: <https://doi.org/10.1109/ICDMW.2013.156>
15. Green D. Episode 16: McKinsey’s Approach to Data-Driven HR (Interview with Keith McNulty, Global Director of People Analytics and Measurement at McKinsey). URL: <https://www.myhrfuture.com/digital-hr-leaders-podcast/2019/10/1/mckinseys-approach-to-data-driven-hr> (accessed: 01.11.2021).
16. Henrys K. Role of Predictive Analytics in Business. 2021. 13 p. URL: <https://ssrn.com/abstract=3829621> (accessed: 01.11.2021).
17. Shen G., Liu Q. Performance analysis of linear regression based on Python. *Communications in Computer and Information Science*, 2020, vol. 1227. DOI: https://doi.org/10.1007/978-981-15-6113-9_80
18. Cherezov D.S., Tjukachev N.A. Overview of the main methods for classifying and clustering data. *Vestnik VGU, Serija: Sistemnyj analiz i informacionnye tehnologii*, 2009, no. 2, pp. 25–29. (In Russ.)
19. SAP BusinessObjects Cloud. URL: <https://www.id-mt.ru/produkty/sap/sap-businessobjects-cloud/> (accessed: 01.11.2021). (In Russ.)
20. SAS Visual Analytics. Tadviser. URL: https://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:SAS_Visual_Analytics (accessed: 01.11.2021). (In Russ.)
21. RAPIDMINER platform for data analysis. Competence Development Center in the field of business informatics of the Graduate School of Business of the National Research University Higher School of Economics. URL: <https://hsbi.hse.ru/articles/platforma-rapidminer-dlya-analiza-dannykh/> (accessed: 01.11.2021). (In Russ.)
22. Alteryx Designer Instruments. URL: <https://biconsult.ru/products/instrumenty-alteryx-designer> (accessed: 01.11.2021). (In Russ.)
23. IBM SPSS applications. URL: <https://www.ibm.com/ru-ru/analytics/spss-statistics-software> (accessed: 01.11.2021). (In Russ.)
24. Microsoft Azure. Tadviser. URL: https://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:Microsoft_Azure (accessed: 01.11.2021). (In Russ.)
25. From January 28, 2021, amendments to the Federal Law «On Advertising» will come into force. URL: <http://kurgan.fas.gov.ru/news/14080> (accessed: 01.11.2021). (In Russ.)
26. Build, train, and deploy machine learning models with a free Azure account. URL: <https://azure.microsoft.com/ru-ru/free/machine-learning/> (accessed: 01.11.2021). (In Russ.)
27. Minka T.P. Bayesian Linear Regression. URL: <https://tminka.github.io/papers/minka-linear.pdf> (accessed: 01.11.2021).
28. Neural network regression. Azure. Machine learning. URL: <https://docs.microsoft.com/ru-ru/azure/machine-learning/algorithm-module-reference/neural-network-regression> (accessed: 01.11.2021). (In Russ.)

29. Bumping Decision Tree Regression Module. Azure. Machine learning. URL: <https://docs.microsoft.com/ru-ru/azure/machine-learning/algorithm-module-reference/boosted-decision-tree-regression> (accessed: 01.11.2021). (In Russ.)
30. Linear regression module. Azure. Machine learning. URL: <https://docs.microsoft.com/ru-ru/azure/machine-learning/algorithm-module-reference/linear-regression> (accessed: 01.11.2021). (In Russ.)
31. Decision forest regression module. Azure. Machine learning. URL: <https://docs.microsoft.com/ru-ru/azure/machine-learning/algorithm-module-reference/decision-forest-regression> (accessed: 01.11.2021). (In Russ.)

Received 08.11.2021

УДК 681.3.06:681

ВЕРОЯТНОСТНЫЕ КРИТЕРИИ ТЕСТОВ ИНФОРМАЦИОННО- ДИАГНОСТИЧЕСКИХ СИСТЕМ

Лихтциндер Б.Я.¹, Ларина В.А.¹, Муравец А.В.²

¹ *Поволжский государственный университет телекоммуникаций и информатики, Самара, РФ*

² *Самарский областной центр медицинской профилактики*

«Центр общественного здоровья», Самара, РФ

E-mail: lixt@psuti.ru

Данная статья посвящена вероятностным критериям тестов информационно-диагностической системы. Рассматриваются вероятностные критерии тестов информационно-диагностической системы, направленной на изучение состояния пациента. Система содержит объект диагностирования (пациент) и средства диагностирования. Система диагностирования реализует некоторый алгоритм диагностирования, задающий последовательность и способ анализа результатов обследования. Целью анализа результатов элементарных проверок является определение диагностируемого состояния, в котором действительно находится пациент. В данной статье рассматриваются такие разделы, как вероятностное диагностирование состояния пациента; матрица условных вероятностей; состояния; шкала состояний; простейшие проверки; условные вероятности. Все данные по каждому разделу сводятся в таблицу для визуализации. Вывод делается на основе анализа, что позволяет оценить эффективность применения данной системы диагностирования.

Ключевые слова: *вероятностные критерии, информационно-диагностические системы, диагностирование состояния пациента, обследование*

Введение

Автоматизированные медико-диагностические системы находят все большее применение при профилактике и диагностировании заболеваний [1; 2; 4; 8–11].

Система диагностирования состояния пациента содержит объект диагностирования ОД (пациент) и средства диагностирования (СД) [3].

СД реализуют некоторый алгоритм диагностирования, задающий последовательность и способ анализа результатов обследования на основе вероятностных критериев проверок [5–7; 12]. Под элементарной проверкой π_k ($k = 1, \dots, n$) будем понимать некоторое тестирующее воздействие, заключающееся в постановке вопроса и анализе ответа на этот вопрос. Под k понимаем последовательное число проверок, прошедших с начала тестирования. (Проверки нумеруются последовательно, по мере их проведения.)

Вероятностное диагностирование состояния пациента

ОД может находиться в одном из диагностируемых состояний e_i на множестве $E = \{e_1, \dots, e_i, \dots, e_S\}$.

Элементами множества E являются состояния e_i ($i = 1, 2, \dots, S$). Имеется набор элементарных проверок $\Pi = \{\pi_1, \dots, \pi_2, \dots, \pi_n\}$.

Целью анализа результатов элементарных проверок является определение диагностируемого состояния, в котором действительно находится ОД.

Каждой проверке π_k соответствует f_k исходов этой проверки, образующих множество $Q_k = \{q_1, \dots, q_{j_k}, \dots, q_{f_k}\}$. Исход q_{j_k} может быть равен единице, если проверка показала результат, соответствующий данному исходу, или равен нулю в противном случае. При каждой проверке