

ПРОГНОЗИРОВАНИЕ ОТГРУЗОК ГОТОВОЙ ПРОДУКЦИИ В РАЗРЕЗЕ КАНАЛОВ СБЫТА НА ОСНОВЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

Куксенюк С.В.

Томский политехнический университет, БШ, аспирант, e-mail: svk82@tpu.ru

Аннотация

Исследовательская работа посвящена проблемам прогнозирования и планирования в коммерческой организации. Цель работы заключается в подборе оптимальных методов прогнозирования объемов продаж производственного предприятия. Результатами исследования, проведенного с использованием возможностей интерактивной среды разработки RStudio, стали рекомендации по повышению эффективности работы аналитического отдела в решении задач прогнозирования объемов продаж и выпуска готовой продукции.

Ключевые слова: анализ временных рядов, прогнозирование, модель Хольта-Уинтерса, линейная регрессия, ARIMA

Введение

Задачи прогнозирования могут быть поставлены в различных областях науки и общественной жизни: от вопросов, касающихся предсказания погоды на ближайшие дни, до попыток предсказать будущие значения медицинских показателей, спрогнозировать факторы риска, течения заболевания для подбора релевантной и своевременной терапии. Однако до сих пор «одной из самых обсуждаемых областей прогнозирования является экономика» [1].

Внешняя среда в наши дни для любого коммерческого предприятия представляется как турбулентная, плохо прогнозируемая, характеризующаяся высоким уровнем изменчивости и неопределенности. Именно поэтому приоритетными для менеджмента современной организации, особенно в условиях распространившегося data-driven подхода для принятия обоснованных управленческих решений, становится организация процессов эффективного прогнозирования и планирования.

Актуальность выбранной темы исследования заключается в ее практической значимости для эффективного ведения бизнеса, повышения качества прогнозов продаж, сокращения убытков и оптимизации процессов, не создающих добавочной стоимости.

Выбор методов прогнозирования

Прежде, чем приступить к описанию используемых в рамках данной работы методов прогнозирования, необходимо описать шаги, которые нам требуется совершить для осуществления анализа и генерации прогнозов, поведения временных рядов в будущем. Для этого выделим шесть этапов, которые были реализованы в ходе нашего исследования:

- 1) подготовка данных;
- 2) визуализация временных рядов;
- 3) реализация моделей и тренировка;
- 4) оценка результативности моделей;
- 5) генерирование прогнозов;
- 6) сравнение, выводы.

В качестве применяемых методов прогнозирования были использованы, как простые методы (среднеарифметический, наивный и наивный с сезонностью, дрейфовый), так и методы линейной регрессии, модели Хольта-Уинтерса с разными типами сезонности, ARIMA. Выбранные модели относятся к группе методов статистического обучения, ставшим «очень популярным во многих научных областях, а также в маркетинге, финансах и других бизнес-дисциплинах» [2].

В ходе выполнения исследовательской работы была осуществлена проверка гипотезы, заключающаяся в предположении, что при сложении результатов прогнозирования, полученных путем отбора лучшей модели для каждого временного ряда (в данном случае обозначающего отгрузки по конкретному каналу сбыта), качество прогнозов по сравнению с использованием аналогичного отбора модели на общем, не детализированном временном ряде, повышается.

Описание датасета

В качестве исходных данных был использован сформированный в базе 1С: Бухгалтерия отчет о продажах девяти SKU одной товарной категории кондитерской продукции с детализацией по контрагентам, месяцам и номенклатуре за период 01.01.2021 г. – 31.12.2023 г. Дальнейшая обработка данных включала в себя процесс сопоставления каждого контрагента с одним из пяти каналов сбыта компании, перевод штучных значений отгрузок в килограммы и консолидацию информации в разрезе каналов продаж по месяцам.

Получившийся датасет получил следующие характеристики: данные структурированы, содержат 3 признака (столбца): channel – категориальная переменная, описывает канал сбыта; date – атрибут даты; value – непрерывная (количественная) переменная, содержащая числа со «знаками после запятой» [3], 180 наблюдений (5 каналов сбыта*36 месяцев наблюдений), пропущенные значения отсутствуют. Данные в нормализации не нуждаются.

Реализация выбранных моделей прогнозирования: тестирование и анализ

Анализ и прогнозирование подготовленных временных рядов выполняется в RStudio средствами библиотеки `frp3`, предназначенной для манипуляций с данными, построения графиков и загружающий другие пакеты, представляющие интерес в целях работы с временными рядами: `dplyr`, `tidyr`, `lubridate`, `ggplot2`.

После установки пакетов для работы с временными рядами конвертируем столбец времени из текста в объект месячного времени, используя функцию `as_tsibble`. После этого проведем визуальный анализ сначала общего временного ряда, отражающего суммарные отгрузки в килограммах по выбранной товарной группе продукции, построим график сезонных подсерий и выполним разложение временного ряда для оценки сезонной составляющей и тренда. Графическое представление перечисленных шагов представлено на рис. 1 и 2.

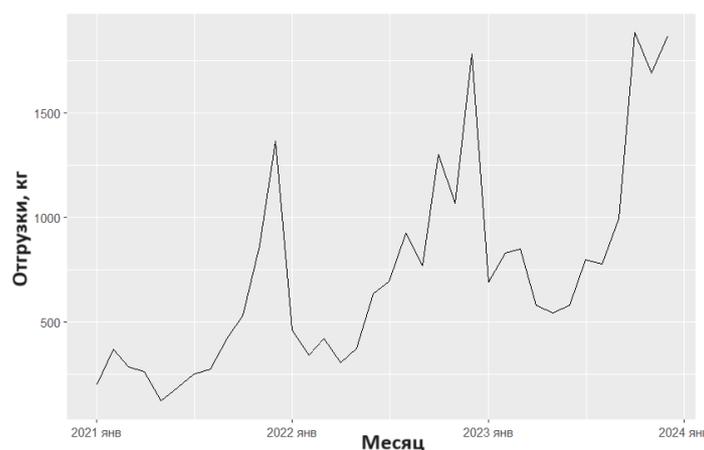


Рис. 1. Совокупные отгрузки по выбранной товарной категории (01.01.2021 г. – 31.12.2023 г.)

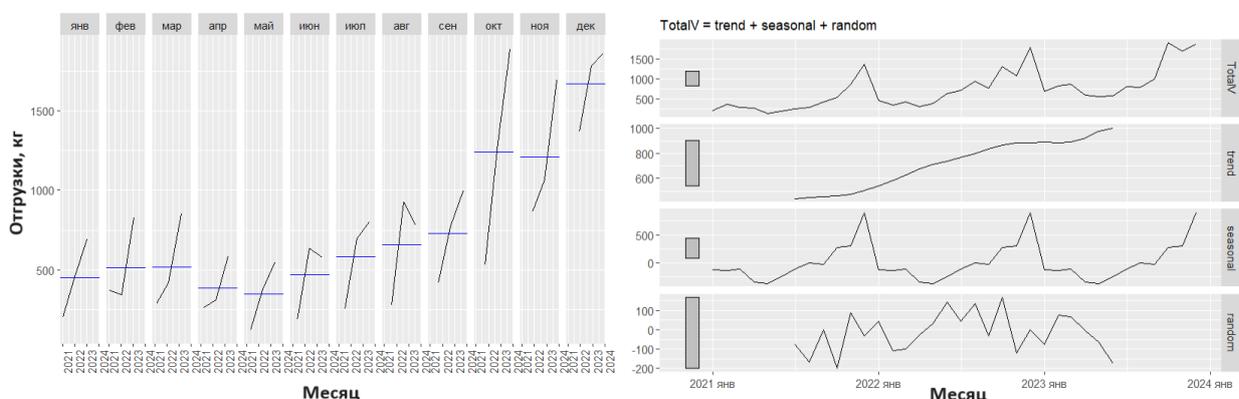


Рис. 2. График сезонных подсерий и классическое аддитивное разложение совокупных отгрузок группы кондитерской продукции

Как видно из графиков, временной ряд содержит явную сезонную составляющую (проявляющуюся значительным увеличением объёмов отгрузок в зимний период (октябрь-декабрь)) и характеризуется наличием восходящего тренда.

Зададим тренировочные данные с января 2021 по май 2023 (80,56 % от всего датафрейма), далее проведем подгонку выбранных моделей, отбор наилучшей модели будем выполнять на основании метрики качества RMSE, обозначающего корень из среднего квадрата ошибки, имеющего «широкое применение на практике» [4] при оценивании точности применяемых методов прогнозирования.

Проведем подбор модели для общего временного ряда. Модели DRIFT, MEAN, NAIVE, Хольта-Уинтерса с мультипликативной сезонностью для усреднения уже исключены, так как показали худшие результаты в первой итерации. Применение нескольких разных методов для одного и того же временного ряда, а затем усреднение результирующих прогнозов – «простой способ повысить точность прогноза» [4]. Помимо прогнозов на 19 месяцев для моделей: наивная с сезонностью, Хольта-Уинтерса с аддитивной сезонностью, TSLM, ARIMA на рис. 3 добавлено использование комбинированного метода, полученного путем усреднения соответствующих значений, прогнозируемых при использовании других моделей. Комбинированный прогноз при этом по большинству мер качества (RMSE, MAE, MASE, RMSSE, ACF1) показывает наилучший результат.

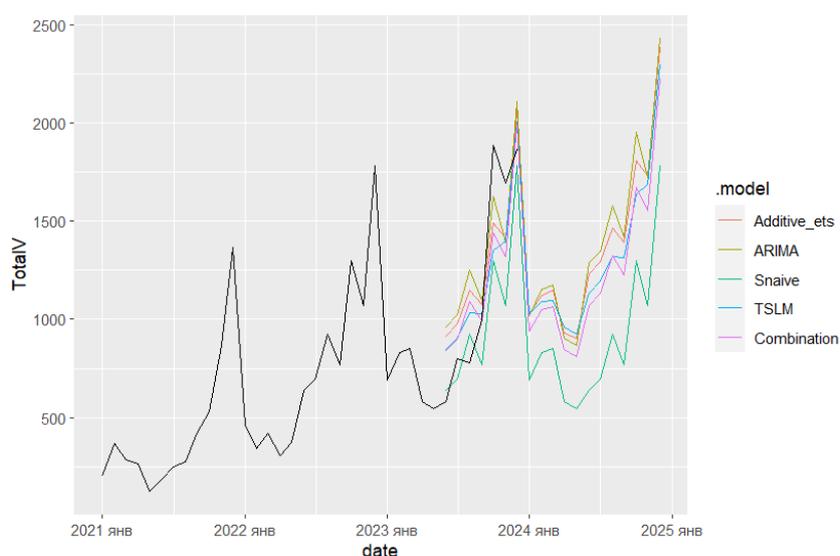


Рис. 3. График с прогнозными значениями по общему временному ряду

Повторим использование отобранных моделей на временных рядах, соответствующих каждому из каналов сбыта, выбрав лучше отработавшие модели по метрике RMSE. Обобщим полученную информацию в табличный вид для удобства интерпретации, указав наименование временного ряда, обозначение модели, показавшей лучший результат по мере качества RMSE и ее значение.

Таблица 1

Отобранные по мере RMSE для генерации прогнозов модели

№	Наименование временного ряда	Модель	RMSE
1	Общий временной ряд	Комбинация прогнозов	274
2	Интернет-магазин	Линейная регрессия	140
3	Коммерческий отдел	Линейная регрессия	108
4	Маркетплейсы	Сезонная наивная	135
5	Спецрозница	Линейная регрессия	22,8
6	Фирменная розница	Комбинация прогнозов	43,3

Теперь используем полученные при применении комбинации прогнозов значения будущих периодов на общем временном ряде и суммируем показатели прогнозов по всем каналам сбыта в разрезе соответствующих месяцев для получения графического отображения двух методов прогнозирования. На рис. 4 исторические данные ограничены временным интервалом тестовой выборки (период с июня 2023 года, 19,44 % от используемого датасета) с наложением прогнозных значений по двум использованным методикам расчета.

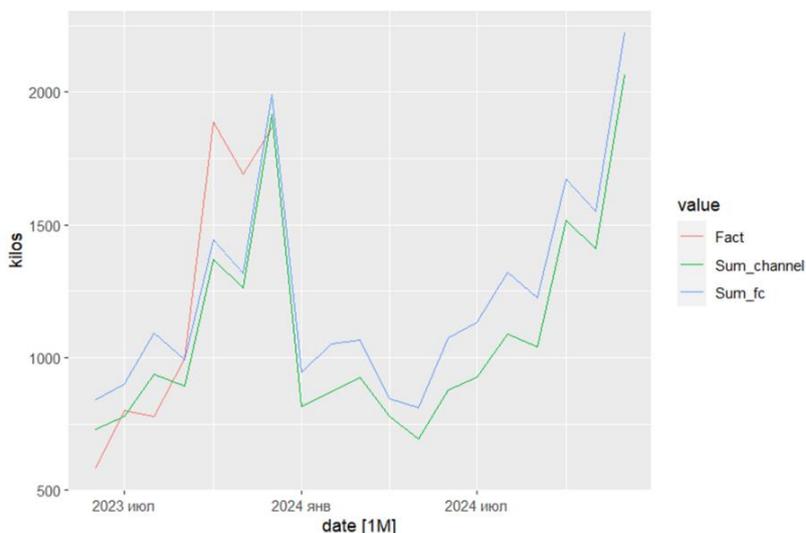


Рис. 4. Графики прогнозных значений, совмещенные с графиком исторических данных. Красным цветом обозначены исторические данные, синим – прогнозные значения по общему временному ряду, зеленым – суммированные прогнозные значения по каналам сбыта

Также становится очевидным, что ни одна из использованных методик не генерирует графиков, повторяющих контуры визуализации исторических данных (пиков сезона). Прогноз на основе применения комбинации методов на общем временном ряде отражает более оптимистичную по отгрузкам картину и в сезонное время чуть ближе к историческим данным, чем при использовании второй методики.

Далее сравним значения полученные при использовании комбинации прогнозов на общем временном ряде; на суммированных значениях; усредненных значениях этих двух прогнозов, данными аналитического отдела с историческими данными по отгрузкам за первые два месяца 2024 года. Критериями для сравнения и ранжирования результатов будут выступать следующие важные для бизнеса приоритеты:

1. Отклонение прогнозируемых значений от факта минимальное.
2. Приоритетнее прогноз с увеличенным объемом (для минимизации недогрузов).
3. Прогнозные значения готовой продукции минимальны.

В соответствии с обозначенными критериями сравним полученные имеющиеся данные.

Таблица 2

Сравнение прогнозов на январь-февраль 2024 года

Месяц	Факт, кг	Усредненные, кг.	Откл-ие, %	Общий, кг.	Откл-ие, %	Каналы, кг.	Откл-ие, %	Аналитик, кг.	Откл-ие, %
январь.24	859	879	2,2 %	942	9,7 %	813	-5,4 %	946	10,1 %
февраль.24	955	960	0,6 %	1050	9,9 %	871	-8,8 %	1431	49,8 %

По обозначенным критериям при сравнении прогнозных значений можно проранжировать источники информации по степени их применимости следующим образом:

1. Комбинация и усреднение сгенерированных нами при помощи RStudio результатов использования двух подходов показали наименьшее отклонение от фактических данных.

2. Прогноз на основе общего временного ряда показывает превышение прогнозных значений от фактического на 9,7 % в январе и 9,9 % в феврале. Учитывая, что увеличенный объем приоритетнее, то применение данного подхода ставим на второе место в нашем списке.

3. Использование подхода суммирования прогнозных значений по каналам сбыта показывает отклонение в меньшую сторону. Использование данного прогноза не обеспечило бы компании необходимым объемом отгрузок, а значит по выбранной товарной группе продукции образовалась бы недополученная прибыль.

4. Прогнозы аналитического отдела показывают худшие результаты, так как их использование привело бы к большому запасу готовой продукции в феврале (на 49,8 %), что характеризуется как критическое превышение поскольку не только создало бы ситуацию «заморозки» в остатках денежных средств, но и угрозу их потери в виду ограниченных сроков годности продукции (условиям контрагентов, предъявляемых к остаточному сроку годности при отгрузке готовой продукции).

Заключение

В продолжении исследования планируется рассмотреть результативность и эффективность применения Prophet, полиномиальной регрессии, нейросетевых моделей. Однако уже на этапе проведенного анализа временных рядов получены результаты, которые могут быть использованы в целях совершенствования, повышения качества процесса прогнозирования, существующего в рассматриваемой коммерческой организации. Наблюдения и закономерности в поведении временных рядов также могут быть использованы менеджментом для принятия управленческих решений: организации оптимального процесса выпуска готовой продукции, оптимизации продолжительности финансового и производственного циклов путем поддержания необходимого для отгрузок запаса товаров на складах.

Список использованных источников

1. Груздев, А.В. Прогнозирование временных рядов с помощью Facebook Prophet, ETNA, sktime и LinkedIn Greykite: Строим, настраиваем, улучшаем модели прогнозирования временных рядов с помощью специальных библиотек [Текст] / А.В. Груздев – М.: ДМК Пресс, 2023. – С. 9
2. Джеймс Г. Введение в статистическое обучение с примерами на языке R. [Текст] / Г. Джеймс, Д. Уиттон, Т. Хасты, Р. Тибширани – М.: ДМК Пресс. – 2016. – С. 11
3. Ын А. Теоретический минимум по Big Data. Всё, что нужно знать о больших данных [Текст] / А. Ын, К. Су – СПб.: Питер, 2022. – С. 25
4. Хайндман Р. Прогнозирование: принципы и практика [Текст] / Р. Хайндман, Д. Атанасопулос – М.: ДМК Пресс, 2023. – С. 149
5. Хайндман Р. Прогнозирование: принципы и практика [Текст] / Р. Хайндман, Д. Атанасопулос – М.: ДМК Пресс, 2023. – С. 427