

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук, Факультет экономических наук  
Образовательная программа «Экономика и анализ данных»

УДК 519.642.4

**Отчет об исследовательском проекте на тему:**  
**Предсказание спроса в ритейле для эффективного управления запасами**  
(промежуточный, этап 1)

**Выполнил студент:**

группы БЭАД231, 2 курса

Чайковский Григорий Тимофеевич

**Принял руководитель проекта:**

Богуцкий Денис Александрович

Проектно-исследовательская лаборатория «ИИ в математических финансах» ФКН

Стажер-исследователь

Москва 2025

## Содержание

<b>Введение</b>	<b>3</b>
<b>Постановка задачи</b>	<b>3</b>
<b>Используемые данные</b>	<b>3</b>
<b>Обзор популярных методов</b>	<b>4</b>
<b>ARIMA</b>	<b>4</b>
<b>VAR</b>	<b>5</b>
<b>Источники</b>	<b>7</b>
<b>Список литературы</b>	<b>7</b>
<b>Приложения</b>	<b>7</b>

## **Введение**

Спецификой ритейла является высокая волатильность спроса и сильно ограниченное время реализации товаров [1]. В этой связи прогнозирование спроса является критически важной задачей для бизнеса, для решения которой используются методы статистики, машинного обучения и нейронных сетей [2].

### **Постановка задачи**

Прогнозирование временных рядов, в частности задача предсказания спроса, является классической задачей регрессии. Требуется найти отображение, которое данному временному ряду - последовательности  $y_1, \dots, y_t$  сопоставит набор предсказаний -  $\bar{y}_{t+1}, \dots, \bar{y}_{t+H}$ , минимизировав разницу между предсказаниями и настоящими значениями  $y_1, \dots, y_{t+H}$

### **Используемые данные**

Проведен описательный анализ публичного датасета о продажах интернет-магазина. Построены графики продаж по каждой категории товара в недельном, месячном и годовом разрезах. Для тестирования предсказательных моделей проведена группировка данных: посчитаны суммы продаж за каждый день.

## Обзор популярных методов

Как уже было сказано, задачу предсказания временных рядов решают разными способами. Во-первых, это «модели временных рядов», предполагающие различные виды линейной зависимости между членами временного ряда. Во-вторых, это методы машинного обучения, такие как линейная регрессия, решающие деревья, kernel regression, градиентный бустинг. Наконец, в последние годы для предсказания временных рядов стали активно использоваться нейрости (artificial neural networks) [1].

### ARIMA

В качестве базовой модели выбрано Auto-regressive integrated moving average, одну из самых популярных статистических моделей для предсказания временных рядов [6]. ARIMA( $p, d, q$ ) предполагает следующую зависимость между членами временного ряда:

$$\begin{aligned} y_t = & \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} \\ & + \beta_1 \varepsilon_1 + \dots + \beta_q \varepsilon_{t-q} + \varepsilon_t \end{aligned} \quad (1)$$

где  $y_t$  - значение в момент  $t$ ,  $\varepsilon_i$  - «белый шум» (компонента ряда, которую невозможно предсказать) в момент  $t$

Для выбора гиперпараметров модели построены графики автокорреляционной и частичной автокорреляционной функций: как известно, оптимальным значением  $p$  в модели AR( $p$ ) считается последний значимый пик ACF, а оптимальным значением  $q$  в MA( $q$ ) - последний значимый пик PACF [5].

Затем процесс подбора гиперпараметров автоматизирован с помощью библиотеки sktime. Для подбора оптимальных значений  $p$ ,  $d$ ,  $q$  использовался grid search - перебор всех разумных комбинаций гиперпараметров - с кросс-валидацией: модель последовательно обучалась на 70 предыдущих значениях и предсказывала 3 следующих. Для наглядности приведен пример кросс-валидации временного ряда, в котором размер тренировочной выборки на каждом шаге равен 5, а размер тестовой так же 3.

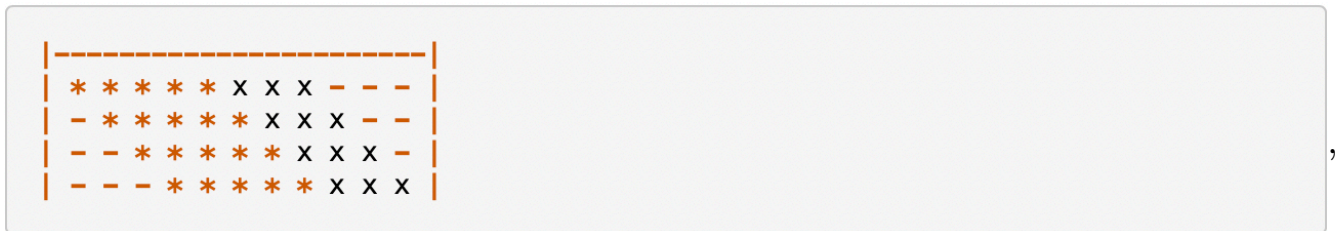


Рис. 1. \* - тренировочные данные, x - тестовые данные

Для сравнения результатов ARIMA с разными наборами  $(p, d, q)$  была выбрана классическая метрика Mean Absolute Percentage Error. Лучшая модель - ARIMA(3, 1, 1) с MAPE=0.31 на тестовой выборке. Таким образом, удалось добиться относительно разумных предсказаний модели.

Проведен анализ ошибок модели: достигнуты требуемые свойства «остатков» временного ряда - некоррелированность, нулевое математическое ожидание. Ошибки имеют распределение, близкое к нормальному.

## VAR

Вместе с тем, кажется разумным, что продажи одних товаров могут влиять на продажи других в разной степени. Как минимум, продажи той же категории товаров в предыдущие периоды влияют на продажи этой категории сегодня больше, чем продажи других категорий. Какие-то товары могут являться компонентами для других: спрос на машины приведет к увеличению спроса на бензин. С другой стороны, возможны и обратные ситуации: повышение спроса на бабл-ти окажет отрицательный эффект на продажи кофе и наоборот, потому что человек скорее всего заменяет один товар другим. Чтобы учитывать эти взаимосвязи, можно применить модель векторной авторегрессии (VAR), что и было сделано. Данная модель также широко применяется для решения поставленной задачи [2, 3].

Данные были разделены на три основных категории, которые торгуются в данном магазине: офисные товары, технологические товары и мебель. Из исходного  $y_t \in \mathbb{R}$  получился  $\bar{y}_t \in \mathbb{R}^3$ , к которому и применялась VAR. Подбор лучший значений был произведен с помощью grid search с использованием многопоточных вычислений для увеличения производительности.

Оказалось, что лучшая модель - это VAR(10). В большинстве случаев удалось добиться неплохих показателей по метрике MAPE, однако из-за того, что в некоторых категориях в некоторые дни продажи могут быть равны нулю, подсчет MAPE

является численно неустойчивым, из-за чего в некоторых случаях получатся экстремально большие значения этой метрики [4].

## **Источники**

### **Список литературы**

- [1] Da Veiga CP, Da Veiga CR, Catapan A, Tortato U, Da Silva WV. Demand forecasting in food retail: A comparison between the Holt-Winters and ARIMA models. WSEAS transactions on business and economics. 2014 Jan;11(1):608-14.
- [2] Tsolacos S. Econometric modelling and forecasting of new retail development. Journal of Property Research. 1998 Jan 1;15(4):265-83.
- [3] Brooks C, Tsolacos S. Forecasting models of retail rents. Environment and Planning A. 2000 Oct;32(10):1825-39.
- [4] Hyndman RJ. Forecasting: principles and practice. OTexts; 2018. URL: <https://otexts.com/fpp2/accuracy.html>
- [5] Robert Nau, Identifying the numbers of AR or MA terms in an ARIMA model, Duke University, 2020 URL: <https://people.duke.edu/~rnau/411arim3.htm>
- [6] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed). Hoboken, New Jersey: John Wiley & Sons.

### **Приложения**

- [1] Репозиторий проекта

URL: [https://github.com/chagrygoris/retail\\_forecasts](https://github.com/chagrygoris/retail_forecasts)