



www.volsu.ru

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В БЕЗОПАСНОСТИ И ТЕЛЕКОММУНИКАЦИЯХ

DOI: <https://doi.org/10.15688/NBIT.jvolsu.2024.2.4>

УДК 004.85

ББК 32.972



АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ СПРОСА НА ТОВАРЫ И УСЛУГИ

Антон Викторович Бондарь

Студент, кафедра информационных систем и компьютерного моделирования,
Волгоградский государственный университет
IVTm-221_731435@volsu.ru
просп. Университетский, 100, 400062 г. Волгоград, Российская Федерация

Александр Сергеевич Астахов

Старший преподаватель, кафедра информационных систем и компьютерного моделирования,
Волгоградский государственный университет
a.s.astahov@volsu.ru
просп. Университетский, 100, 400062 г. Волгоград, Российская Федерация

Аннотация. В статье рассмотрены основные алгоритмы для прогнозирования временных рядов методами машинного обучения, в частности прогнозирование продаж товаров на основе различных показателей. Задача прогнозирования требует от персонала отличного владения математическими и статистическими инструментами, а также умение анализировать большие объемы данных. Автоматизация данной задачи позволит переложить большую часть работы сотрудников на программное обеспечение. Это поможет увеличить объем обрабатываемой информации, снизить затрат на логистику и хранение, а также минимизировать риски потери прибыли из-за нулевого остатка на складе. В статье проведен анализ критериев, влияющих на спрос товаров; рассмотрены классические алгоритмы и нейронные сети для прогнозирования временных рядом. Также в работе освещен процесс проектирования, разработки, тестирования программного обеспечения.

Ключевые слова: временные ряды, прогнозирование временных рядов, линейная регрессия, ARIMA, XGBoost, нейронные сети, рекуррентные нейронные сети, LSTM, веб-технологии.

**Подготовка и анализ набора данных.
Критерии оценки качества
прогнозирования**

Временной ряд – набор данных, отсортированный по временной составляющей и характеризующий состояние элементов в конкретные периоды времени [4]. Для набора данных была получена и подготовлена информация о продажах товаров из разных категорий за период с 2021 по 2023 г. из официального API маркетплейса Wildberries. В качестве наиболее значимых признаков, влияющих на продажи, были выделены следующие:

- 1) габариты и размерность товара;
- 2) категория товара;
- 3) продажи за прошлые периоды;
- 4) цена с учетом скидки и размер скидки;
- 5) номер дня в году и в неделе;
- 6) конверсия карточки товара из кликов в корзину и из кликов в продажи.

Для оценки качества прогнозирования использовались следующие метрики:

1. R^2 – коэффициент детерминированности, имеющий вид

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

где y_i значения прогнозируемой переменной; n – количество наблюдений в выборке; $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$ – среднее значение по экспериментальным данным; \tilde{y}_i – прогнозируемые значения [2].

2. MAE – средняя абсолютная ошибка, имеющая вид

$$MAE = \frac{\sum_{i=1}^n |y_i - \tilde{y}_i|}{n},$$

где n – количество наблюдений в выборке; \tilde{y}_i – прогнозируемые значения; y_i – фактические значения искомой переменной [2].

3. MSE – среднеквадратичная ошибка

$$MAE = \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{n},$$

где n – количество наблюдений в выборке; \tilde{y}_i – прогнозируемые значения; y_i – фактические значения искомой переменной [2].

**Классические методы
прогнозирования временных рядов**

Линейная регрессия является статистической моделью, которая предполагает, что прогнозируемая переменная линейно зависит от одной или нескольких независимых переменных [9]. Основная идея линейной регрессии заключается в поиске линейной функции, которая наилучшим образом соответствует наблюдаемым данным. Математически, линейная регрессия имеет вид:

$$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_n * x_n + \varepsilon,$$

где x_1, x_2, \dots, x_n – независимые переменные; $\beta_1, \beta_2, \dots, \beta_n$ – коэффициенты регрессии; ε – ошибка модели. Значения метрик получились следующие [9]:

1. $R^2 = 0.91$.
2. $MAE = 277$.
3. $MSE = 193169$.

График фактических и прогнозных значений представлен на рисунке 1.

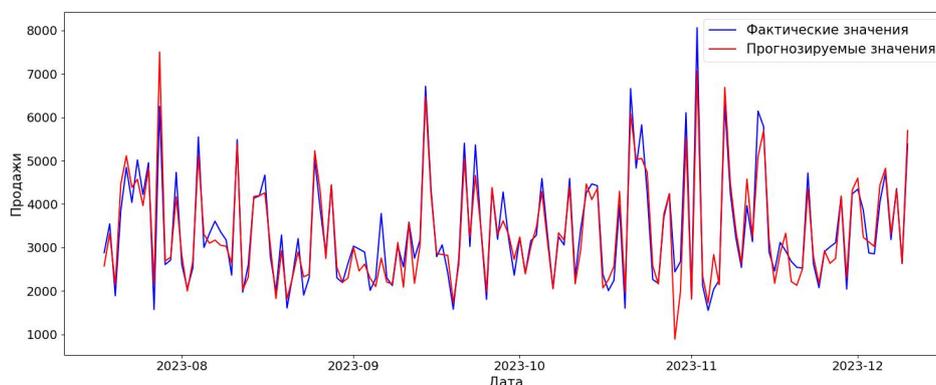


Рис. 1. График фактических и спрогнозированных значений, полученный с помощью линейной регрессии

ARIMA (Autoregressive Integrated Moving Average) – алгоритм основанный на комбинации трех этапов: авторегрессии, интеграции и скользящего среднего [1]. График полученных значений представлен на рисунке 2. Значения метрик получились следующие:

1. $R^2 = 0.93$.
2. $MAE = 140$.
3. $MSE = 53613$.

XGBoost (Extreme Gradient Boosting) – алгоритм градиентного бустинга, разработанный для решения задач классификации и регрессии. Он основан на принципе ансамбля моделей, где каждая модель стремится исправить ошибки предыдущей модели [3]. В процессе обучения XGBoost строит деревья решений последовательно, учитывая градиенты потерь от предыдущих моделей.

При проверке качества прогнозирования был сделан вывод, что модель недостаточно

качественно описывает вариации в данных, также имеет высокие среднюю абсолютную ошибку и среднеквадратичную ошибку, что видно на рисунке 3, а также по следующим показателям:

1. $R^2 = 0.83$.
2. $MAE = 362$.
3. $MSE = 345369$.

Еще одним методом прогнозирования временных рядов являются рекуррентные нейронные сети – сети, в основе которых лежит последовательный анализ и обработка информации, с учетом данных из предыдущего вычисления [5; 7]. Схема рекуррентной нейронной сети представлена на рисунке 4.

LSTM является разновидностью рекуррентных нейронных сетей, способная запоминать и использовать информацию на протяжении длительных временных интервалов [6; 8].

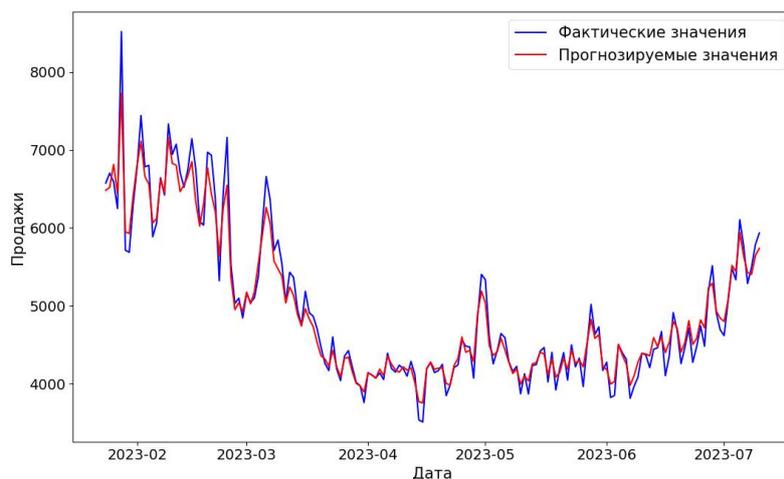


Рис. 2. График фактических и спрогнозированных значений, полученный с помощью модели ARIMA

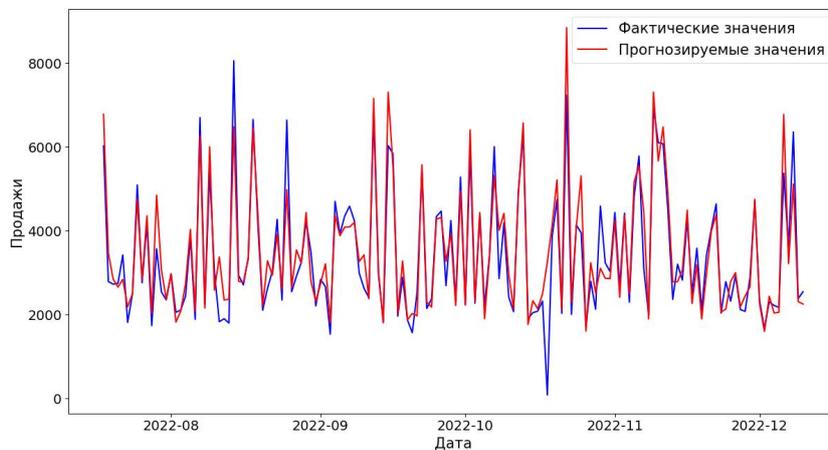


Рис. 3. График фактических и спрогнозированных значений, полученный с помощью XGBoost

Проверка на качество и точность прогнозирования показала хорошие результаты:

1. $R2 = 0.89$.
2. $MAE = 409$.
3. $MSE = 370950$.

График полученных значений представлен на рисунке 5.

Проектирование и разработка программного обеспечения

Вначале была построена логическая структура веб-сервиса на которой видно все

необходимые страницы. После того, как определились с функциями, которые должен выполнять веб-сервис, и его макетом, была спроектирована база данных, представленная на рисунке 6.

Итоговый результат блока для прогнозирования продаж представлен на рисунке 7.

Чтобы воспользоваться прогнозированием продаж, необходимо зайти на детальную страницу товара, выбрать количество дней для прогноза и запустить расчет. Ниже появится таблица с прогнозными значениями, полученными разными моделями прогнозирования.

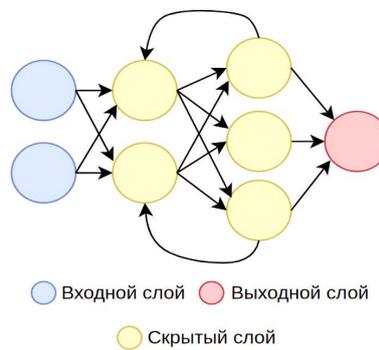


Рис. 4. Схема рекуррентной нейронной сети

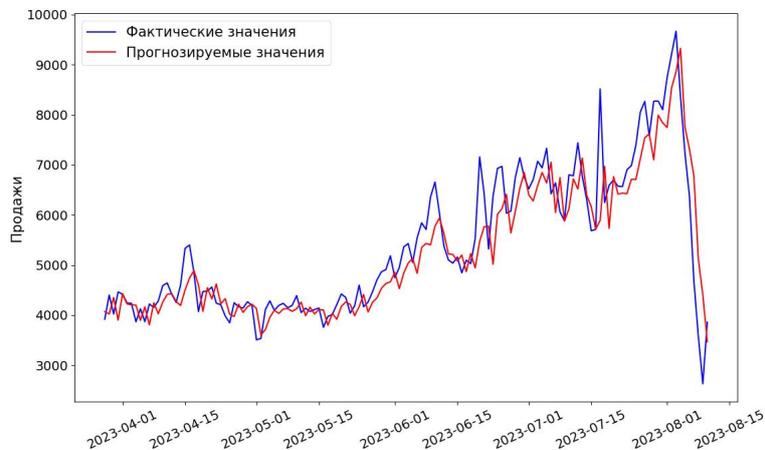


Рис. 5. График фактических и спрогнозированных значений, полученный с помощью LSTM



Рис. 6. Схема базы данных

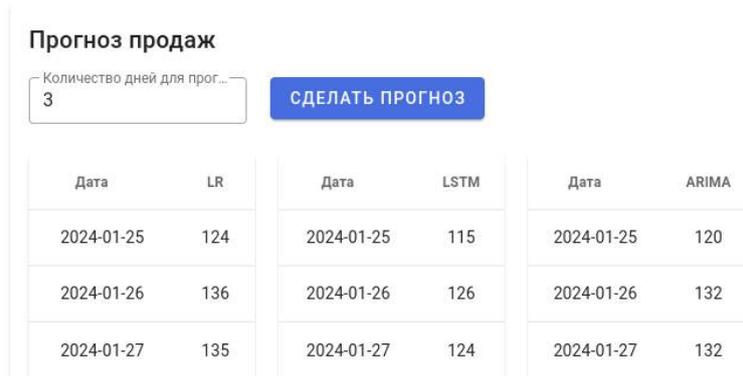


Рис. 7. Блок прогнозирования продаж

Заключение

В результате работы был разработан веб-сервис, позволяющий прогнозировать продажи по товарам на Wildberries. Были рассмотрены классические методы: линейная регрессия, ARIMA, XGBoost. Также реализована и протестирована модель рекуррентной нейронной сети LSTM. В процессе работы, был собран и обработан набор данных о продажах с маркетплейса Wildberries за 2020–2023 годы. Для получения данных использовалось официальное API Wildberries. Результаты данного исследования показали, что лучше всего с задачей справляются модели линейной регрессии и ARIMA с точностью предсказания 91 % и 93 % соответственно. Модель нейронной сети LSTM показала результат несколько хуже классических методов – 89 %. В результате работы было обнаружено, что методами машинного обучения возможно добиться хороших результатов в прогнозировании спроса на товары и услуги, и итоговое программное обеспечение полностью выполняет поставленные задачи.

СПИСОК ЛИТЕРАТУРЫ

1. Афанасьев, В. Н. Анализ временных рядов и прогнозирование / В. Н. Афанасьев, М. М. Юзбашев. – М. : Финансы и статистика, 2010. – 320 с.
2. Голубев, В. В. Введение в математическую статистику / В. В. Голубев. – М. : Моск. физико-техн. ин-т, 2014. – 127 с.
3. Макаров, Д. А. Использование алгоритма XGBoost для предсказания завершения курса обу-

чающимся / Д. А. Макаров // StudNet. – 2021. – Т. 4, № 1. – С. 18.

4. Плотников, А. Н. Элементарная теория анализа и статистическое моделирование временных рядов / А. Н. Плотников. – СПб. : Лань, 2016. – 218 с.

5. Medsker, L. R. Recurrent Neural Networks. Design and Applications / L. R. Medsker, L. C. Jain. – Boca Raton : CRC Press, 2001. – 389 p.

6. Nurfaizi, A. Ticket Prediction Using LSTM on a GLPI System / A. Nurfaizi, M. Hasanuddin // International Journal of Open Information Technologies. – 2023. – No. 7. – P. 118–121.

7. Radivojac, P. Machine Learning Handbook / P. Radivojac, M. White. – 2019. – 140 p.

8. Sanchez, E. N. Discrete-Time Recurrent Neural Control: Analysis and Applications / E. N. Sanchez. – Boca Raton : CRC Press, 2022. – 292 p. – DOI: 10.1201/9781315147185

9. Tsitsiashvili, G. Calculating the Variance of the Linear Regression Coefficient / G. Tsitsiashvili // Reliability: Theory & Applications. – 2019. – Vol. 14, No. 3 (54). – P. 65–68. – DOI: 10.24411/1932-2321-2019-13008

REFERENCES

1. Afanasyev V.N., Yuzbashev M.M. *Analiz vremennyh ryadov i prognozirovanie* [Time Series Analysis and Forecasting]. Moscow, Finansy i statistika Publ., 2010. 320 p.

2. Golubev V.V. *Vvedenie v matematicheskuyu statistiku* [Introduction to Mathematical Statistics]. Moscow, Mosk. fiziko-tehn. in-t, 2014. 127 p.

3. Makarov D.A. *Ispolzovanie algoritma XGBoost dlja predskazaniya zaversheniya kursa obuchajushhimsja* [Using the XGBoost Algorithm to Predict the Completion of the Course for Students]. StudNet, 2021, vol. 4, no. 1, p. 18.

4. Plotnikov A.N. *Elementarnaja teorija analiza i statisticheskoe modelirovanie vremennyh rjadov* [Elementary Theory of Analysis and Statistical Modeling of Time Series]. Lan Publ., 2016. 218 p.

5. Medsker L.R., Jain L.C. *Recurrent Neural Networks. Design and Applications*. CRC Press, 2001. 389 p.

6. Nurfaizi A., Hasanuddin M. Ticket Prediction Using LSTM on a GLPI System. *International Journal of Open Information Technologies*, 2023, vol. 11, no. 7, pp. 118-121.

7. Radivojac P., White M. *Machine Learning Handbook*, 2019. 140 p.

8. Sanchez E.N. *Discrete-Time Recurrent Neural Control: Analysis and Applications*. CRC Press, 2022. 292 p. DOI: 10.1201/9781315147185

9. Tsitsiashvili G. Calculating the Variance of the Linear Regression Coefficient / G. Tsitsiashvili. *Reliability: Theory & Applications*, 2019, vol. 14, no. 3 (54), pp. 65-68. DOI 10.24411/1932-2321-2019-13008

MACHINE LEARNING ALGORITHMS FOR FORECASTING DEMAND FOR GOODS AND SERVICE

Anton V. Bondar

Student, Department of Information Systems and Computer Modeling,
Volgograd State University
IVTm-221_731435@volsu.ru
Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Alexander S. Astakhov

Senior Lecturer, Department of Information Systems and Computer Modeling,
Volgograd State University
a.s.astahov@volsu.ru
Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Abstract. The article discusses the main algorithms for forecasting time series using machine learning methods, in particular, forecasting sales of goods based on various indicators. The forecasting task requires the staff to have excellent knowledge of mathematical and statistical tools, as well as the ability to analyze large amounts of data. Automating this task will allow you to shift most of the work of employees to software. This will help to increase the volume of information processed, reduce logistics and storage costs, and minimize the risks of loss of profit because of a zero stock balance. The article analyzes the criteria that affect the demand for goods; classical algorithms and neural networks for forecasting time series are considered. The work also highlights the process of designing, developing, and testing software.

Key words: time series, time series forecasting, linear regression, ARIMA, XGBoost, neural networks, recurrent neural networks, LSTM, web technologies.