

development of a promising enterprise.countries. It is concluded that the enterprises of the oil and gas industry were able to cope with difficulties and maintain their positions.

Keywords: content management, modeling, information system, database design.

About the authors:

Kalach Andrey Vladimirovich – Doctor of Chemical Sciences, Professor, Head of the Department of Information Security and Protection of State Secret Information, Voronezh Institute of the Federal Penitentiary Service of Russia, Voronezh. E-mail: a_kalach@mail.ru

Danilova Victoria Alekseevna – Cadet, Voronezh Institute of the Federal Penitentiary Service of Russia, Voronezh. E-mail: viktory.danilova2021@mail.ru

УДК 004.85, 658.8

ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ ИНСТРУМЕНТОВ ПРЕДИКТИВНОЙ АНАЛИТИКИ ДЛЯ АНАЛИЗА ДЕЯТЕЛЬНОСТИ ПРЕДПРИЯТИЯ

Н.В. Корнеева, Е.И. Корнеева

© Корнеева Н.В., Корнеева Е.И., 2024

Аннотация. В статье рассмотрена предиктивная аналитика и ее практическое применение для анализа деятельности предприятия. Задача прогнозирования данных решается с помощью системы Loginom Community и пакета Anaconda.

Ключевые слова: Loginom Community, Anaconda, Python, предиктивная аналитика, большие данные, гребневая регрессия, временные ряды, SARIMAX.

Цифровые технологии активно внедряются в анализ деятельности предприятия. В распоряжении Правительства РФ [1] указано, что данные в цифровой форме являются ключевым фактором производства во всех сферах социально-экономической деятельности. Объем данных растет в геометрической прогрессии (463 эксабайта данных к 2025 г. [2]), и их эффективная обработка возможна только при наличии развитых платформ и технологий. «Большие данные» используются для своевременной оценки рисков, автоматизации рутинных бизнес-процессов компании [3, 4].

Аналитика данных понимается как процесс поиска системных закономерностей в массивах информации и их интерпретации с целью получения важных для бизнеса сведений. Можно выделить четыре вида аналитики данных: описательную (дескриптивную), диагностическую, предиктивную (предсказательную), предписывающую. Дескриптивная аналитика используется для первичного анализа данных и рассматривает текущую ситуацию. Предиктивная прогнозирует события по исходному набору данных, в том числе с применением методов машинного обучения (machine learning) и интеллектуального анализа данных (data mining) [4]. Методы машинного обучения рассматриваются как алгоритмы, которые обучаются на размеченных цифровых данных и после выгрузки моделей работают без вмешательства человека на основе выработанных шаблонов и логических выводов [5].

Анализ данных предприятия и построение прогноза реализуются с помощью цифровых платформ и языков программирования. Для предиктивной аналитики необходимо использовать данные предприятия. В исследовании использовались следующие файлы из открытых данных компании 1С [6]:

sales_train.csv – данные продаж для обучения алгоритмов за два года;

test.csv – данные продаж для тестирования алгоритмов;

items.csv – файл с информацией о товарах;

item_categories.csv – файл с информацией о возможных категориях товаров;

shops.csv – файл с информацией о магазинах компании.

Таблицы на ресурсе Kaggle содержат около трех миллионов строк.

Анализ данных с платформой Loginom Community

Работа в Loginom Community представляет собой последовательное добавление узлов и выполнение настроек элементов. Версия Community устанавливается после регистрации на локальный компьютер [7, 8]. Общий вид схемы для исходных данных представлен на рис. 1.

Задача предсказания продаж по времени решена с помощью узла ARIMAX, или модели временных рядов (рис. 1). Модель учитывает влияние прошлых значений переменной, влияние ошибок прогнозов на прошлые периоды, автоматизирует процесс прогнозирования с минимальным количеством входных данных (зависимость одного поля от временного столбца).

Для алгоритмов регрессии также требуется группировка данных в случаях, когда данные содержат множество категориальных переменных или когда оценивается влияние различных категорий на зависимую переменную.

В схему добавлен узел «Линейная регрессия» с настройками ядра для гребневой регрессии (ridge regression). Такой вариант настройки включает регуляризацию линейной регрессии и помогает предотвратить переобучение модели путем добавления штрафа к размеру коэффициентов регрессии. Модель построена как зависимость числовых столбцов набора данных компании 1С от столбца с суммой продаж. Время транзакции не учитывается.

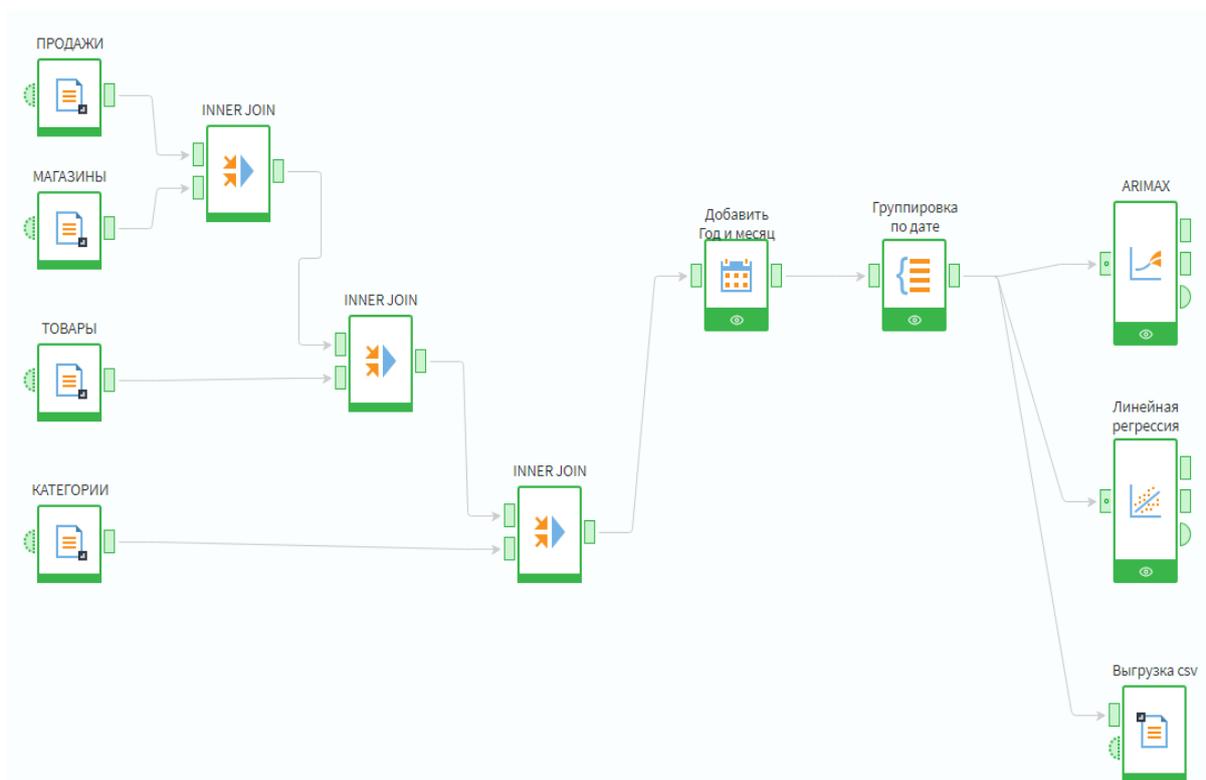


Рис. 1. Сценарий обработки данных продаж компании 1С в программе Loginom Community (выполнено авторами)

Узлы ARIMAX и «Линейная регрессия» в Loginom автоматически подсчитывают качество построенной модели по коэффициенту детерминации R^2 и по другим критериям. Раздел «Визуализация узла» позволяет построить график прогноза модели, достаточно настроить характеристики данных, которые отображаются по оси абсцисс и ординат.

Анализ данных с Jupyter Notebook и Anaconda

Установка и обновление Jupyter Notebook и других компонентов Anaconda может быть сложной задачей, требующей определенного уровня знаний и опыта.

Прежде чем начинать анализ данных, необходимо проверить, что Jupyter Notebook запускается на локальном компьютере и настроены дополнительные библиотеки Python [9]. Для исследования необходимо загрузить исходные данные и объединить их в общую таблицу с помощью библиотеки pandas. В Loginom Community числовые данные и даты выбраны в узле группировка, остальные столбцы отфильтрованы средствами low-code платформы. Отличие настройки в Anaconda состоит в том, что общая схема реализуется не визуально, а последовательными командами на языке Python в ячейках Jupyter Notebook.

Исходные данные [6] включают также столбцы с текстовой информацией. После слияния всех файлов, добавления столбца с месяцем и годом покупки и группировки по месяцу и году необходимо убрать из данных для моделей наименования магазинов и другие текстовые столбцы. После подготовки данных в Jupyter Notebook добавлены ячейки с методами ARIMAX и SARIMAX, а также гребневой регрессией (библиотеки auto_arima, statsmodels, модули linear_model и pipeline библиотеки scikit-learn) [10]. Для того чтобы оценить качество моделей, применяются метрики качества: коэффициент детерминации и информационные критерии Байеса и Акаике, загружаемые вручную из раздела metrics библиотеки scikit-learn [10]. Визуализация строится отдельно с помощью настроек библиотеки matplotlib.

Оценка качества моделей

Модели временных рядов после ряда экспериментов были выбраны по минимальным критериям: AIC-критерий (Акаике) равен 302,56 и BIC-критерий (Баейса) равен 303,13. С помощью моделей спрогнозированы продажи на ближайший год и на более длительный срок с учетом сезонности (рис. 2).

Модели регрессии построены с высоким R^2 (коэффициент детерминации), равным 0,88, и низкими значениями ошибок регрессии MSE, RMSE, MAE как в Loginom Community, так и в Anaconda. На рис. 3 представлен автоматический отчет с критериями качества регрессии в Loginom.

Таким образом, результаты анализа для исходных данных компании 1С удовлетворительны по критериям до 2015 г. (открытые данные в источнике [6]). Для завершения отчета и выводов по предиктивной аналитике требуется уточнение и проведение дополнительных испытаний.

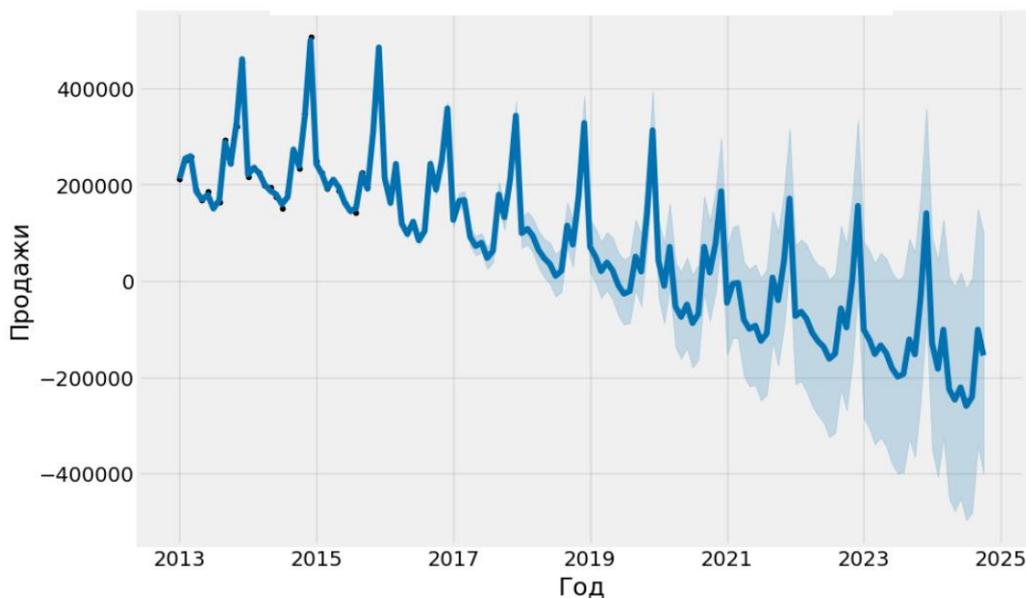


Рис. 2. Прогноз продаж компании 1С по модели временных рядов SARIMAX до октября 2024 года (выполнено авторами)

Линейная регрессия • Быстрый просмотр					
Выход регрессии		Коэффициенты регрессионной модели		Сводка	
№	Имя	Метка	Значение		
1	12 TotalSamples	Всего примеров	34		
2	12 TotalSelectedSamples	Всего отобранных примеров	34		
3	12 TrainSamples	Примеров в обучающем множестве	24		
4	9.0 LogLikelihood	Логарифм функции правдоподобия	-266,78		
5	9.0 R2	Коэффициент детерминации	0,88		
6	9.0 AdjustedR2	Скорректированный коэффициент детерминации	0,86		
7	9.0 StdDev	Стандартное отклонение	4 871 187,13		
8	12 DFE	Число степеней свободы ошибки	12		
9	12 ModelDF	Число степеней свободы модели	3		
10	9.0 FStatistic	F-статистика	30,56		
11	9.0 AIC	Информационный критерий Акаике	33,85		
12	9.0 AICc	Информационный критерий Акаике скорректированный	34,08		
13	9.0 BIC	Информационный критерий Байеса	34,04		
14	9.0 HQC	Информационный критерий Ханнана-Куинна	33,86		
15	9.0 ModelPValue	P-значение модели	0,00		
16	12 TestSamples	Примеров в тестовом множестве	10		
17	12 ValidationSamples	Примеров в валидационном множестве	5		
18	12 ValidationPartitions	Число валидационных разбиений	5		

Рис. 3. Сводная оценка показателей регрессии в Loginom Community (выполнено авторами)

Loginom Community и Anaconda используются для решения задач в области предиктивной аналитики. Loginom обеспечивает упрощение

процесса обработки данных, автоматическую оценку качества алгоритмов и для стандартных задач не требует написания кода на Python и SQL (low-code). С другой стороны, Anaconda – фреймворк, для применения которого необходимы базовые знания в области программирования и который обладает более широким функционалом, чем Loginom (1500 библиотек). Обе технологии используются для прогнозирования в различных областях экономики на основе анализа больших данных и машинном обучении.

Библиографический список

1. Программа «Цифровая экономика Российской Федерации»: Распоряжение Правительства РФ от 28 июля 2017 г. № 1632-р. URL: <https://base.garant.ru/71734878/> (дата обращения: 10.04.2024).

2. Морроу Дж. Как вытащить из данных максимум. Навыки аналитики для неспециалистов. М.: Альпина Паблицер, 2022. С. 15–62.

3. Кузнецова Н.В., Харитонов А.А. Бизнес-анализ: ключевые концепции, методы и преимущества // Baikal Research Journal. 2023. № 1. URL: <https://cyberleninka.ru/article/n/biznes-analiz-klyuchevye-kontseptsii-metody-i-preimuschestva> (дата обращения: 10.04.2024).

4. Цыгулева А.А. Анализ и структуризация поля знаний методов предиктивной аналитики // Теория и практика современной науки. 2022. № 12 (90). URL: <https://cyberleninka.ru/article/n/analiz-i-strukturizatsiya-polyaznaniy-metodov-prediktivnoy-analitiki> (дата обращения: 10.04.2024).

5. Sarker I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions // SN COMPUT, 2021. SCI. 2. URL: <https://doi.org/10.1007/s42979-021-00592-x> (дата обращения: 12.04.2024).

6. Predict Future Sales. Daily sales of 1C Company competition for Coursera course // Kaggle. URL: <https://www.kaggle.com/c/competitive-data-science-predict-future-sales/overview> (дата обращения: 16.04.2024).

7. Loginom – руководство пользователя // Loginom Company URL: <https://help.loginom.ru/userguide/index.html> (дата обращения: 16.04.2024).

8. Волков Н. Модели вида ARIMA: учебник по машинному обучению. URL: <https://education.yandex.ru/handbook/ml/article/modeli-vida-arima> (дата обращения: 17.04.2024).

9. User guide // Anaconda, Inc. URL: <https://docs.conda.io/projects/conda/en/latest/user-guide/index.html> (дата обращения: 17.04.2024).

10. Scikit-learn. Machine Learning in Python // Scikit-learn. URL: <https://scikit-learn.org/stable/index.html> (дата обращения: 19.04.2024).

Об авторах:

Корнеева Наталья Викторовна – старший преподаватель кафедры экономики и управления производством, ФГБОУ ВО «Тверской государственный технический университет», Тверь. E-mail: nvkorneewa@yandex.ru

Корнеева Елена Игоревна – старший преподаватель кафедры программного обеспечения, ФГБОУ ВО «Тверской государственный технический университет», Тверь. E-mail: yelena.korneeva@yandex.ru

PREDICTIVE ANALYTICS TOOLS: A PRACTICAL APPROACH TO BUSINESS ANALYSIS

N.V. Korneeva, E.I. Korneeva

Abstract. The article discusses predictive analytics and its practical application in analyzing the activities of a business. The task of forecasting data is solved using the Loginom Community platform and the Anaconda package.

Keywords: Loginom Community, Anaconda, python, predictive analytics, big data, ridge regression, time series, SARIMAX.

About the authors:

Korneeva Natalia Victorovna – Senior Lecturer of the Department of Economy and Production Management Department, Tver State Technical University, Tver. E-mail: nvkorneewa@yandex.ru

Korneeva Elena Igorevna – Senior Lecturer of the Department of Software Department, Tver State Technical University, Tver. E-mail: yelena.korneeva@yandex.ru

УДК 004.451:004.9

ОСНОВНЫЕ ОФИСНЫЕ ПРОГРАММНЫЕ ПАКЕТЫ, ПОДДЕРЖИВАЕМЫЕ ОПЕРАЦИОННОЙ СИСТЕМОЙ ASTRA LINUX

Г.В. Кошкина, К.Э. Никитина-Кошкина, Е.Е. Фомина

© Кошкина Г.В., Никитина-Кошкина К.Э.,
Фомина Е.Е., 2024

Аннотация. В статье выполнено сравнение основных возможностей пакетов офисных программ LibreOffice, R7-Офис, МойОфис. Описаны их компоненты и возможности. Кратко рассмотрена операционная система Astra Linux, ее возможности и недостатки.

Ключевые слова: операционная система, Astra Linux, LibreOffice, R7-Офис, МойОфис, программные пакеты.