

УДК 519.246.8, 004.9

Обзор open-source библиотек для решения задач прогнозирования временных рядов

Свекольников Е.А.*

Московский авиационный институт
(национальный исследовательский университет) (МАИ)
г. Москва, Российская Федерация
ORCID: <https://orcid.org/0009-0000-6161-571X>
e-mail: elena.cvekolnikova@gmail.com

Пановский В.Н.**

Московский авиационный институт
(национальный исследовательский университет) (МАИ)
г. Москва, Российская Федерация
ORCID: <https://orcid.org/0009-0007-1708-8984>
e-mail: panovskiy.v@yandex.ru

Представлен обзор разнообразных open-source Python-библиотек, предназначенных для анализа и прогнозирования временных рядов. Охватываются такие инструменты, как Prophet, Kats, Merlion, а также алгоритмы ARIMA, LSTM, позволяющие исследовать сезонность, тренды и аномалии в данных временных рядов. Подробно рассмотрены возможности каждой библиотеки, их преимущества и области применения в анализе временных данных.

Ключевые слова: Python-библиотека, временные ряды, open source, прогнозирование, тренд, Prophet, Kats, Merlion.

Для цитаты:

Свекольников Е.А., Пановский В.Н. Обзор open-source библиотек для решения задач прогнозирования временных рядов // Моделирование и анализ данных. 2024. Том 14. № 2. С. 45–61. DOI: <https://doi.org/10.17759/mda.2024140203>

*Свекольников Елена Андреевна, студентка, кафедра «Математическая кибернетика», Московский авиационный институт (национальный исследовательский университет) (МАИ), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0009-0000-6161-571X>, e-mail: elena.cvekolnikova@gmail.com

**Пановский Валентин Николаевич, кандидат физико-математических наук, доцент кафедры «Математическая кибернетика», Московский авиационный институт (национальный исследовательский университет) (МАИ), г. Москва, Российская Федерация, ORCID: <https://orcid.org/0009-0007-1708-8984>, e-mail: panovskiy.v@yandex.ru



1. ВВЕДЕНИЕ

В современной жизни во многих сферах трудно обойтись без временных рядов. Временные ряды применяются для хранения вполне ожидаемых последовательностей данных, например, сведений об изменениях складских запасов и погодных условий [1]. Это может быть любая информация, измеряемая через определенные промежутки времени: дни, месяцы, часы, минуты и т.д.

Временные ряды широко используются в различных областях, включая авиацию, финансы, экономику, метеорологию, медицину, техническое обслуживание и другие [1, 2]. Анализ временных рядов позволяет обнаруживать тренды, сезонные колебания, циклы и другие закономерности, что помогает в принятии решений, выявлении паттернов и прогнозировании данных. Прогнозирование – это важная часть инженерного процесса: предсказать предел выносливости деталей [17], показатели качества, надежности [18] и другие [19], количество остаточных ресурсов [20].

Изучение и анализ временных рядов имеет большое значение для прогнозирования будущих значений, выявления закономерностей, внедрения эффективных стратегий и принятия обоснованных решений на основе исторических данных. По сути, временные ряды являются необходимым инструментом во многих научных областях.

Предсказание временных рядов играет важную роль в различных областях кибернетики и информационных технологий, где анализ временных данных и прогнозирование будущих значений имеют ключевое значение. Вот некоторые области, где предсказание временных рядов чрезвычайно важно:

1. Финансовые рынки [1]:
 - Прогнозирование цен на финансовых рынках для принятия решений о торговле и инвестициях.
 - Моделирование волатильности цен для рискованного управления.
2. Промышленность и производство [1]:
 - Прогнозирование спроса на продукцию для планирования производства и управления запасами.
 - Предсказание отказов оборудования для предупреждения и проведения обслуживания.
3. Климатология и метеорология: предсказание погоды и климатических изменений для планирования действий в экстренных ситуациях и адаптации к изменениям климата [1].
4. Транспорт и логистика [2]:
 - Прогнозирование транспортного спроса для оптимизации маршрутов и планирования логистики.
 - Предсказание трафика для улучшения управления дорожным движением.
5. Здравоохранение [1]:
 - Прогнозирование распространения для разработки и реализации мер по предотвращению заболеваний и эпидемий и их контролю.
 - Предсказание медицинских показателей для диагностики заболеваний и планирования лечения.



6. Маркетинг и реклама: прогнозирование поведения потребителей для оптимизации маркетинговых стратегий и увеличения эффективности рекламы [1].

Анализ временных рядов и предсказание их значений являются важными инструментами в различных областях кибернетики и информационных технологий для прогнозирования трендов, принятия решений и оптимизации бизнес-процессов.

2. ОБЗОР OPEN-SOURCE БИБЛИОТЕК

В современном мире важно уметь пользоваться открытыми и доступными инструментами, чтобы прежде всего уметь задать базовый уровень для создания новых алгоритмов. К таким инструментам относятся:

- Google Colab (облачная среда для работы с Python, удобна для анализа данных и временных рядов) [6],
- Kaggle (платформа для соревнований по анализу данных, предоставляет доступ к множеству данных для работы с временными рядами) [7],
- библиотеки pandas, NumPy [8], Matplotlib [9], Seaborn [10], Statsmodels [11], Prophet [12], Kats [13], Merlion [14] в Python,
- Jupyter Notebook: популярная среда для интерактивных вычислений.

Эти инструменты предоставляют мощные средства для анализа и прогнозирования временных рядов, визуализации данных и создания моделей временных рядов.

3. PROPHET

Остановимся подробнее на библиотеках. Prophet – это библиотека для анализа и прогнозирования временных рядов. Её основные особенности и преимущества:

- Простота в использовании: Prophet обладает простым и интуитивно понятным программным интерфейсом приложения, что делает его доступным для широкого круга пользователей даже без глубоких знаний в анализе временных рядов.
- Устойчивость к отсутствию данных: может автоматически обрабатывать отсутствующие данные и аномалии, что упрощает процесс работы с реальными временными данными. Аномалия, или выброс, определяется как элемент, явно выделяющийся из набора данных, к которому он принадлежит, и существенно отличающийся от других элементов выборки [5].
- Учет сезонности и праздников: библиотека позволяет учитывать сложные сезонные и праздничные эффекты, что особенно важно для прогнозирования временных рядов с сезонными колебаниями.
- Гибкость в настройке: предлагает пользователю возможность настройки модели на основе его потребностей, включая изменение сезонности, добавление праздников и определение структуры тренда.
- Интерпретируемые результаты: Prophet генерирует прогнозы с интерпретируемым разложением на тренд, сезонность и праздники, что облегчает анализ и понимание прогнозов.



- Открытый исходный код: библиотека является проектом с открытым исходным кодом, что способствует активному развитию и поддержке сообществом пользователей.
- Визуализация позволяет делать интуитивно понятные графики. На графике изображено, как тренд и сезонность взаимодействуют и формируют общую динамику временного ряда (рис. 1).

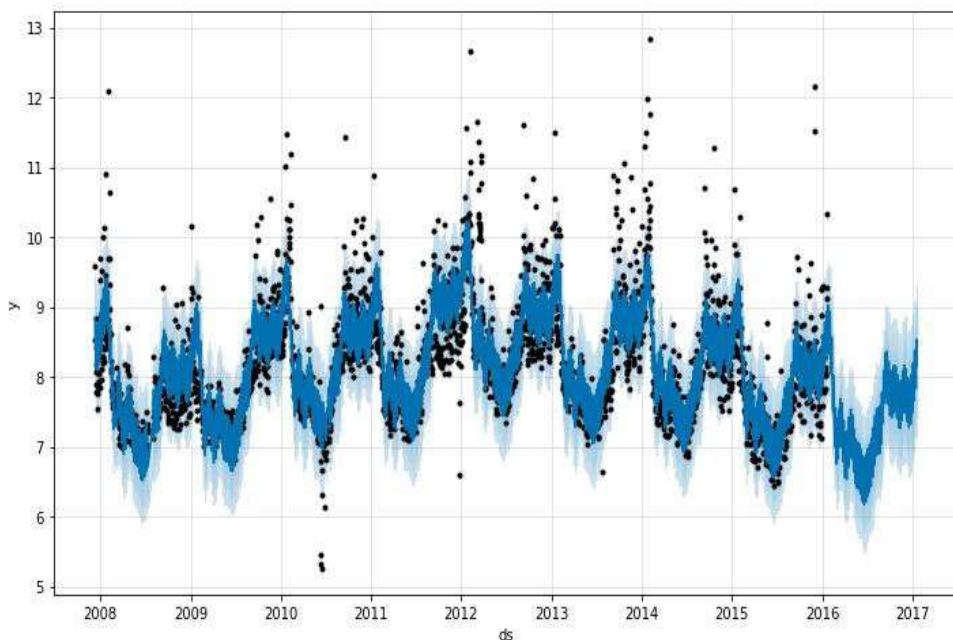


Рис. 1. Визуализация тренда и сезонности в Prophet

Благодаря простоте использования, гибкости в настройке и возможности учета сложных сезонных эффектов, Prophet стал популярным инструментом для анализа и прогнозирования временных рядов в различных областях.

В библиотеке Prophet в основном используется адаптивный алгоритм анализа временных рядов, основанный на аддитивной регрессии «trend + seasonality + holidays» (тренд + сезонность + праздники).

Основные компоненты модели в Prophet включают [12]:

1. Тренд (Trend):
 - Prophet автоматически определяет и моделирует общий тренд изменений в данных временного ряда.
 - Тренд может быть линейным или нелинейным, в зависимости от структуры данных.
2. Сезонность (Seasonality):
 - Prophet учитывает сезонные колебания в данных, например, ежедневные, еженедельные, ежемесячные или годовые паттерны.



- Модель Prophet может учесть как аддитивные, так и мультипликативные компоненты сезонности.
- 3. Праздники (Holidays):
 - Библиотека Prophet предоставляет возможности для включения праздников и важных событий в модель.
 - Праздничные дни могут оказывать значительное влияние на временной ряд.
- 4. Шум и аномалии: Prophet умеет обрабатывать аномалии и шум в данных, предоставляя возможности для их учета при анализе временных рядов.

Этот набор компонентов позволяет модели Prophet адаптироваться к различным структурам и особенностям временных рядов, обеспечивая устойчивость к отсутствию данных, изменчивости в данных и корректировку прогнозов в соответствии с обнаруженными аномалиями и праздничными эффектами.

4. KATS

Другой библиотекой в Python, используемой для работы с временными рядами, является Kats. Kats (Kit to Analyze Time Series) – это мощная библиотека для анализа временных рядов. Она предоставляет широкий спектр функций и алгоритмов для работы с временными данными.

Основные возможности Kats:

1. Моделирование и прогнозирование: различные алгоритмы, такие как ARIMA, LSTM (Long Short-Term Memory), TBATS (Trigonometric seasonal components, Box-Cox transformation, ARMA errors, Trend and Seasonal components) и другие, для моделирования и прогнозирования временных рядов.
2. Детекция аномалий: Kats включает инструменты для обнаружения аномалий и нестандартных паттернов во временных рядах.
3. Преобразование данных: имеет функции для подготовки и обработки временных данных, включая работу с пропущенными значениями.
4. Анализ и визуализация:
 - Визуализация К-мер спектров помогает визуально исследовать уникальные шаблоны в данных (рис. 2).
 - Обладает возможностями для визуализации данных и результатов моделирования.
 - Возможность проводить анализ сезонности, трендов и других характеристик временных рядов.

Преимущества использования Kats:

- Многофункциональность: предлагает разнообразные инструменты для полноценного анализа временных рядов.
- Простота использования: обладает удобным интерфейсом и интуитивно понятными функциями.

Kats – это надежный инструмент для анализа временных рядов, который объединяет в себе множество функций, начиная от моделирования и прогнозирования



и заканчивая анализом и визуализацией данных. Kats может стать отличным выбором для решения задач анализа и прогнозирования.

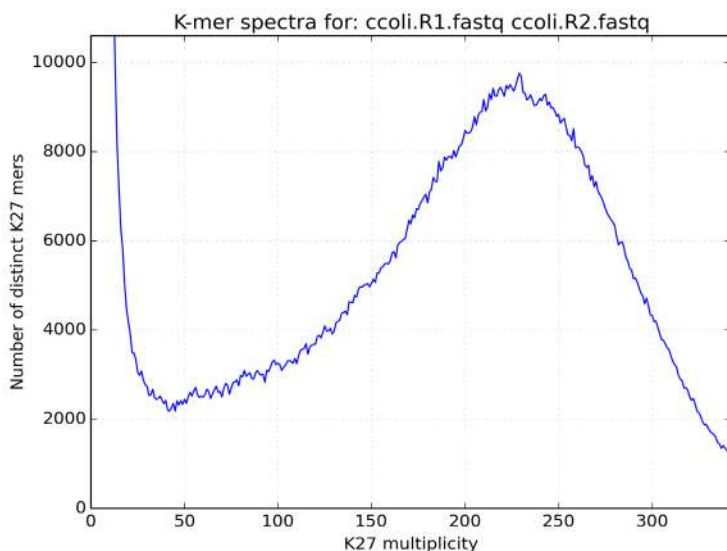


Рис. 2. Визуализация данных в Kats

В библиотеке Kats представлены разнообразные алгоритмы для работы с временными рядами:

1. Prophet-ARIMA: комбинация модели Prophet и модели ARIMA для улучшения точности прогнозирования.
2. Arima: модель ARIMA (Autoregressive Integrated Moving Average) для анализа и прогнозирования временных рядов.
3. LSTM: рекуррентная нейронная сеть, применяемая для моделирования временных данных с учетом долгосрочных зависимостей.
4. TBATS: комбинированный метод, учитывающий сезонность, тренды и другие компоненты временных рядов.

Это лишь небольшой список ключевых алгоритмов, доступных в Kats. Библиотека Kats также включает другие методы и инструменты для работы с временными рядами, такие как детекция аномалий, визуализация данных, подготовка данных и другие функции, которые помогут в анализе и прогнозировании временных данных.

5. TBATS

TBATS (Trigonometric seasonal components, Box-Cox transformation, ARMA errors, Trend and Seasonal components) – это комплексный метод анализа и прогнозирования временных рядов, который учитывает множество компонентов и особенностей временных данных для создания достоверных прогнозов.

Основные особенности:

1. Учет сезонности: обеспечивает учет различных видов сезонности в данных, таких как дневная, недельная, месячная сезонности и другие.
2. Применение преобразований Вох-Сох: используются для стабилизации дисперсии и приведения данных к более нормальному распределению.
3. Учет авторегрессионных ошибок ARMA: включает в себя компоненты авторегрессии и скользящего среднего для обработки корреляции ошибок в данных.
4. Моделирование трендов и сезонности: TBATS моделирует как долгосрочные тренды, так и сезонные колебания временных рядов для точного прогнозирования.
5. Гибкость и настройка: позволяет настраивать параметры модели в зависимости от особенностей конкретного временного ряда для достижения оптимальных результатов.

Преимущества использования метода:

- Универсальность: способность моделировать различные сезонности и компоненты временных рядов.
- Прогностическая точность: обеспечивает более точные прогнозы благодаря учету нескольких аспектов данных.
- Гибкость настройки: позволяет адаптировать модель к специфическим характеристикам конкретного временного ряда.

Метод TBATS представляет собой инструмент для анализа и прогнозирования временных рядов, объединяя в себе различные аспекты, такие, как сезонность, тренды, авторегрессии и другие компоненты. Его гибкость и способность к адаптации делают его привлекательным выбором для повышения точности и достоверности прогнозов в различных областях анализа временных данных.

6. LSTM

Другим инструментом в Kats, используемым для анализа временных рядов, является LSTM:

LSTM (Long Short-Term Memory) – это вид рекуррентных нейронных сетей, специально разработанный для обработки последовательных данных и умения запоминать долгосрочные зависимости во временных рядах.

Основные особенности:

1. Долгосрочная память: способны эффективно удерживать информацию на протяжении длительных периодов времени, что делает их особенно полезными для анализа временных рядов с долгосрочными зависимостями.
2. Избегание проблем затухания (градиенты становятся слишком малыми по мере распространения через глубокие слои нейронной сети, поэтому веса слоев не обновляются, и сеть перестает обучаться) и взрыва градиентов (градиенты становятся очень большими, из-за чего обновления весов становятся больше чем требуется, что может сбить оптимизацию сети).



3. Гибкость моделирования: могут быть настроены для моделирования различных видов временных рядов, а также включать дополнительные компоненты, такие как сезонность или тренды.
4. Адаптивная сеть: способность LSTM к адаптации к изменяющимся условиям временных данных позволяет им эффективно прогнозировать и анализировать временные ряды даже в условиях изменчивости.

Преимущества использования:

- Умение запоминать зависимости в данных: могут удерживать информацию на длительные временные интервалы, что позволяет им учитывать сложные временные зависимости.
- Гибкость моделирования: возможность настройки под различные характеристики временного ряда делает их удобным инструментом анализа данных.
- Применение в различных областях: LSTM успешно применяются в прогнозировании рядов, обработке текстов, обработке изображений и других областях машинного обучения.

Метод представляет собой эффективный и гибкий подход к анализу временных рядов, способный учитывать сложные структуры и зависимости в данных. Их способность к долгосрочному запоминанию и адаптации делает LSTM одним из наиболее мощных инструментов в области анализа временных данных.

7. ARIMA

В Kats есть модель ARIMA, которую активно применяют для работы с временными рядами:

ARIMA (Autoregressive Integrated Moving Average) – это модель временных рядов, которая объединяет компоненты авторегрессии (AR - Autoregressive) и скользящего среднего (MA - Moving Average) с интегрированием (I-Integrated), что позволяет эффективно моделировать и прогнозировать временные данные.

Основные характеристики:

1. Авторегрессия: отвечает за моделирование зависимости текущего значения временного ряда от предыдущих значений в этом же ряду.
2. Интегрирование: позволяет обработать нестационарность данных путем их дифференцирования до достижения стационарности. Нестационарность данных означает непостоянность статистических характеристик ряда со временем.
3. Скользящее среднее (будущие значения предсказываются по известным предыдущим значениям [1]): учитывает зависимость текущего значения ряда от ранее случившихся ошибок модели.

Применение:

- ARIMA в Kats позволяет анализировать временные ряды, моделировать тренды и сезонные колебания, а также совершать прогнозирование на основе исторических данных.
- Он эффективен для прогнозирования временных рядов с линейными зависимостями и относительно стабильными трендами.

Преимущества использования:

- Простота применения: ARIMA позволяет легко моделировать временные ряды без большого числа настраиваемых параметров.
- Устойчивость к стационарности: применение интегрирования в ARIMA позволяет учесть случаи, когда временные данные нестационарны.

ARIMA широко используется для анализа и моделирования временных рядов, особенно в случаях, когда данные обладают линейной структурой и стационарностью. В контексте Kats, ARIMA открывает возможности для прогнозирования временных рядов, что делает его ценным инструментом для исследования временных данных.

8. MERLION

Merlion – это библиотека для анализа временных рядов, разработанная Яндексом. Она предлагает разнообразные инструменты и методы для работы с временными данными.

Работа с временными рядами в Merlion:

1. Загрузка данных: предоставляет удобные способы загрузки и подготовки временных данных из различных источников, включая CSV-файлы, базы данных и другие форматы.
2. Предобработка и очистка данных: позволяет проводить предварительную обработку данных, включая удаление пропущенных значений, фильтрацию выбросов и приведение данных к нужному формату.
3. Моделирование и прогнозирование: предлагает широкий спектр моделей для анализа и прогнозирования временных рядов, включая SARIMA, ARIMA, Prophet, LSTM и другие.
4. Оценка качества модели: позволяет оценить качество построенных моделей с использованием различных метрик, таких как среднеквадратичная ошибка (MSE), коэффициент детерминации и других [1].
5. Разработка собственных моделей: помимо встроенных алгоритмов можно разрабатывать собственные модели для анализа временных рядов, что расширяет возможности анализа данных.
6. Визуализация: Merlion обладает возможностями для визуализации временных рядов, прогнозов моделей, а также показателей качества предсказаний, что способствует более наглядному анализу данных. Позволяет обнаружить аномалии с помощью модели DefaultDetector (рис. 3).

Преимущества использования:

- Широкий выбор моделей: разнообразные методы и алгоритмы для работы с временными рядами.
- Удобство в использовании: интуитивный интерфейс, что упрощает анализ и моделирование временных данных.
- Гибкость и масштабируемость: способен эффективно обрабатывать как небольшие, так и крупномасштабные временные ряды.

Merlion обладает разнообразными функциональностями для работы с данными, позволяет исследовать, моделировать и прогнозировать временные данные, делая процесс анализа более эффективным и удобным.

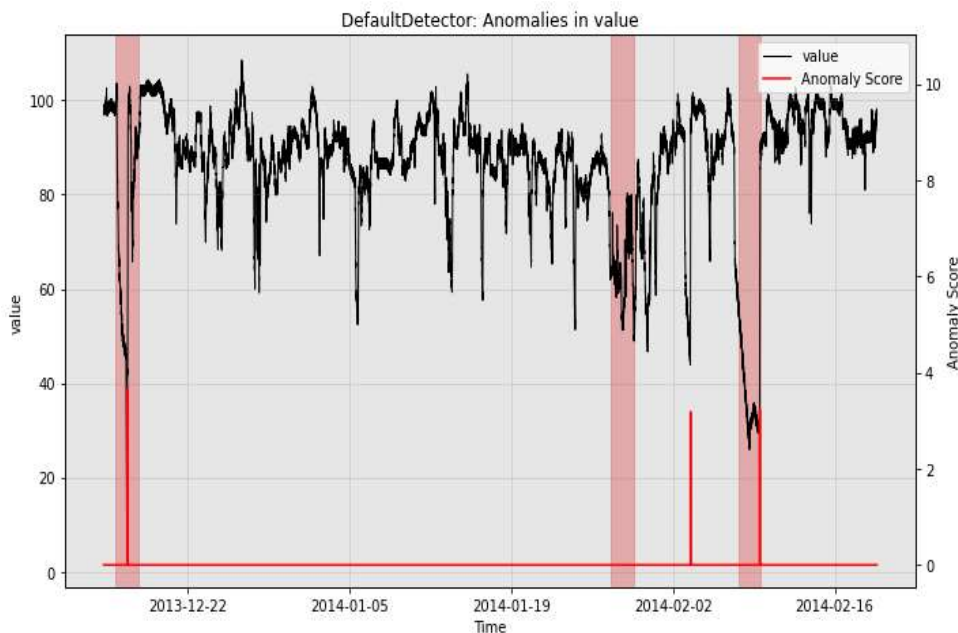


Рис. 3. Визуализация аномалий в Merlion

В библиотеке Merlion для анализа временных рядов доступны различные алгоритмы и модели:

1. ARIMA (Autoregressive Integrated Moving Average): модель для моделирования и прогнозирования временных рядов, учитывающая авторегрессию, интегрирование и скользящее среднее.
2. SARIMA (Seasonal Autoregressive Integrated Moving-Average): расширение модели ARIMA с учетом сезонности для анализа и прогнозирования временных данных.
3. Prophet: модель, разработанная для прогнозирования временных рядов с учетом сезонности и праздников.
4. LSTM (Long Short-Term Memory): рекуррентная нейронная сеть LSTM для работы с динамическими временными данными и изучения долгосрочных зависимостей.
5. Классическая экспоненциальная сглаживающая модель (Exponential Smoothing): модель для прогнозирования временных рядов, учитывающая экспоненциально убывающие веса для прошлых значений.
6. DeepAR: модель DeepAR для прогнозирования с учетом автокорреляции и нелинейных зависимостей в данных [15].



7. Модели обнаружения аномалий: Merlion также предлагает алгоритмы для обнаружения аномалий во временных рядах, что помогает выявлять необычные и выбивающиеся паттерны.

Это лишь небольшой перечень некоторых доступных алгоритмов в библиотеке Merlion: она предлагает широкий спектр методов и моделей для анализа, моделирования и прогнозирования временных рядов, делая ее универсальным инструментом для работы с данными временных рядов.

9. SARIMA

В библиотеке Merlion есть модель SARIMA (Seasonal Autoregressive Integrated Moving-Average), которая является расширением модели ARIMA для учета сезонности во временных рядах. Она включает в себя компоненты авторегрессии, интегрирования, скользящего среднего и учитывает сезонный компонент для более точного прогнозирования временных данных.

Основные характеристики:

1. Сезонность: позволяет учитывать сезонные колебания во временных рядах, что делает ее эффективной для прогнозирования данных с периодическими изменениями.
2. Стационарность: учитывает нестационарность временных данных через интегрирование, что позволяет обеспечить адекватное моделирование и прогнозирование.
3. Авторегрессия и скользящее среднее: присутствие компонентов авторегрессии и скользящего среднего позволяет учесть зависимость текущих значений ряда от предыдущих значений и ошибок модели.

Применение модели:

- Для анализа и прогнозирования временных рядов с учетом сезонных изменений и долгосрочных зависимостей.
- Может быть эффективно применена для данных, в которых присутствует ярко выраженная сезонность.

Преимущества использования:

- Учет сезонности: позволяет учесть цикличность и регулярные изменения во временных рядах.
- Прогнозирование точных значений: модель особенно полезна для точного прогнозирования в данных с сезонной динамикой.
- Универсальность: SARIMA является гибкой моделью, которая может быть настроена по различным сценариям анализа временных данных.

Способность учитывать сезонные компоненты и долгосрочные зависимости делает SARIMA важным инструментом для работы с данными, где присутствует периодичность изменений.



10. EXPONENTIAL SMOOTHING

Классическая экспоненциальная сглаживающая модель (Exponential Smoothing) – это метод прогнозирования временных рядов, основанный на усреднении предыдущих значений с использованием экспоненциального взвешивания, чтобы учесть влияние более недавних данных на прогноз.

Основные характеристики:

1. Взвешивание данных: экспоненциальное сглаживание позволяет придать больший вес более свежим данным, что делает модель более чувствительной к последним наблюдениям.
2. Простота и эффективность: метод является простым в реализации и обладает хорошей способностью адаптироваться к изменчивости временных данных.
3. Учет трендов и сезонности: в классической модели экспоненциального сглаживания можно встретить вариации, учитывающие тренды и сезонность в данных.

Применение метода:

- Прогнозирование временных рядов: используется для прогнозирования будущих значений временных рядов на основе предыдущих данных.
- Анализ трендов и структуры данных: этот метод помогает выявить тренды и основные паттерны во временных данных.

Преимущества использования:

- Простота и понятность: экспоненциальное сглаживание просто в использовании и понимании.
- Хорошая адаптивность: работает с быстро меняющимися данными и обеспечивает гибкую настройку.
- Эффективность в обработке больших объемов данных: справляется с большим объемом временных данных и быстро вычисляет прогнозы.

Классическая экспоненциальная сглаживающая модель в Merlion представляет собой простой, но многофункциональный метод для анализа и прогнозирования временных рядов. Ее преимущества включают простоту использования, адаптивность к изменчивым данным и хорошую способность к прогнозированию основных паттернов в данных.

11. ЗАКЛЮЧЕНИЕ

Прогнозирование временных рядов применяется в авиационной и ракетно-космической технике для планирования предупредительного технического обслуживания или замены деталей до возможного отказа, позволяет прогнозировать процесс износа материалов и компонентов, расход топлива, траектории ракет и спутников, пассажиропоток, загруженность аэропортов.

Благодаря прогнозированию временных рядов появляется возможность улучшить управление рисками, повысить безопасность полетов, оптимизировать процессы и ресурсы, а также повысить эффективность систем в целом.



В мире анализа временных рядов Python представляет множество мощных библиотек, которые обеспечивают широкие возможности для анализа, моделирования и прогнозирования данных. Такие библиотеки, как Prophet, Kats, Merlion и другие предоставляют спектр методов и инструментов, позволяющих исследовать сезонность, тренды, аномалии и другие характеристики временных рядов.

Python-библиотеки обладают гибкостью, простотой использования и возможностью настройки под различные типы данных, что предоставляет широкий выбор для исследователей, аналитиков и специалистов по анализу данных. Сочетание разнообразных алгоритмов, визуализации и анализа данных делает эти библиотеки необходимыми инструментами для успешного решения задач, связанных с временными рядами. Внедрение этих инструментов в рабочий процесс поможет сделать анализ временных данных более точным, информативным и эффективным. В табл. 1 представлены результаты сравнения библиотек и методов, рассмотренных в статье.

Таблица 1

Сравнение библиотек и их методов

Библиотека/метод	Основные преимущества
Prophet	Автоматически обрабатывает многие характерные особенности временных рядов, такие как изменчивость тренда, сезонность и праздники.
Kats	Многофункциональность: предлагает разнообразные инструменты для полноценного анализа временных рядов (TBATS, LSTM, ARIMA).
TBATS	Способность моделировать временные ряды с несколькими сложными компонентами, такими как сезонность, тренды, изменяющаяся дисперсия и авторегрессионная составляющая ошибок.
LSTM	Возможность эффективно удерживать информацию на протяжении длительных периодов времени, что делает их особенно полезными для анализа временных рядов с долгосрочными зависимостями.
ARIMA	Моделирование сложных временных рядов, учитывая как авторегрессионные компоненты (зависимость от предыдущих значений временного ряда), так и скользящие средние компоненты (управление случайной изменчивостью данных).
Merlion	Разработка собственных моделей для анализа и прогнозирования временных рядов, что расширяет возможности анализа данных.
SARIMA	Позволяет моделировать сложные сезонные зависимости в данных, что делает его идеальным инструментом для прогнозирования временных рядов с явно выраженной сезонностью.
Exponential Smoothing	Подходит для прогнозирования временных рядов с быстро меняющимися трендами и колебаниями: за счет весов, присваиваемых различным точкам данных, в зависимости от их удаленности от текущего момента времени.



Литература

1. *Эйлин Нильсен*. Практический анализ временных рядов: прогнозирование со статистикой и машинное обучение.: Пер. с англ. – СПб.: ООО “Диалектика”, 2021. – 544 с.
2. *Бокс Дж., Дженкинс Г.М.* Анализ временных рядов, прогноз и управление. М.: Мир, 1974. 406 с.
3. *Дж. Вандер Плас*. Python для сложных задач: наука о данных и машинное обучение. – СПб.: Питер, 2018. – 576 с.
4. *Чернобровов Алексей*. ТОП-7 Python-библиотек для анализа временных рядов. URL: <https://chernobrovov.ru/articles/top-7-python-bibliotek-dlya-vremennyh-ryadov.html>. (Дата обращения: 02.03.2024)
5. *С.Г. Антипов, М.В. Фомина*. Проблема обнаружения аномалий в наборах временных рядов. Программные продукты и системы, № 2 (98) 2012 г. – 168с. – С. 78–82.
6. Google Colab: [Электронный ресурс]. URL: <https://colab.research.google.com/>. (Дата обращения: 04.03.2024)
7. Kaggle: [Электронный ресурс]. URL: <https://www.kaggle.com/>. (Дата обращения: 04.03.2024)
8. NumPy: [Электронный ресурс]. URL: <https://numpy.org/doc/>. (Дата обращения: 04.03.2024)
9. Matplotlib: [Электронный ресурс]. URL: <https://matplotlib.org/stable/index.html>. (Дата обращения: 04.03.2024)
10. Seaborn: [Электронный ресурс]. URL: <https://seaborn.pydata.org/>. (Дата обращения: 04.03.2024)
11. Statsmodels: [Электронный ресурс]. URL: <https://www.statsmodels.org/stable/index.html>. (Дата обращения: 04.03.2024)
12. Prophet: [Электронный ресурс]. URL: <https://prophet.readthedocs.io/en/latest/>. (Дата обращения: 04.03.2024)
13. Kats: [Электронный ресурс]. URL: <https://kat.readthedocs.io/en/latest/>. (Дата обращения: 04.03.2024)
14. Merlion: [Электронный ресурс]. URL: <https://opensource.salesforce.com/Merlion/latest/index.html>. (Дата обращения: 04.03.2024)
15. DeepAR Forecasting Algorithm: [Электронный ресурс]. URL: <https://docs.aws.amazon.com/sagemaker/latest/dg/deepar.html>. (Дата обращения: 04.03.2024)
16. Python: [Электронный ресурс]. URL: <https://www.python.org/> (Дата обращения: 04.03.2024)
17. *Письмаров А.В.* К вопросу прогнозирования предела выносливости резьбовых деталей с поверхностным упрочнением // Труды МАИ. 2023. № 129. URL: <https://trudymai.ru/published.php?ID=173003>
18. *Пуганов М.Н., Куликов А.В., Новомейский Д.Н.* Прогнозные математические модели тонкоплёночных элементов микросборок // Труды МАИ. 2023. № 131. URL: <https://trudymai.ru/published.php?ID=175920>
19. *Блинов А.В., Разумов Д.А.* Процедура формализации стратегий как элемент методики учета факторов неопределенности при прогнозировании показателей



- реализации программ создания космической техники // Труды МАИ. 2022. № 122. URL: <https://trudymai.ru/published.php?ID=164270>
20. *Беляев Б.В., Лебедев А.С.* Методика прогнозирования остаточного ресурса при разгерметизации летательных аппаратов // Труды МАИ. 2022. № 125. URL: <https://trudymai.ru/published.php?ID=168167>



Review of Open-Source Libraries for Solving Time Series Forecasting Problems

Elena A. Svekolkova*

Moscow Aviation Institute (National Research University) (MAI), Moscow, Russia

ORCID: <https://orcid.org/0009-0000-6161-571X>

e-mail: elena.cvekolnikova@gmail.com

Valentin N. Panovskiy**

Moscow Aviation Institute (National Research University) (MAI), Moscow, Russia

ORCID: <https://orcid.org/0009-0007-1708-8984>

e-mail: panovskiy.v@yandex.ru

An overview of various open-source Python libraries for time series analysis and forecasting is presented. It covers such tools as Prophet, Kats, Merlion, as well as ARIMA, LSTM algorithms, which allow to study seasonality, trends and anomalies in time series data. The capabilities of each library, their advantages and applications in time series data analysis are discussed in detail.

Keywords: Python library, time series, open source, forecasting, trend, Prophet, Kats, Merlion.

For citation:

Svekolkova E.A., Panovskiy V.N. Review of Open-Source Libraries for Solving Time Series Forecasting Problems. *Modelirovanie i analiz dannykh = Modelling and Data Analysis*, 2024. Vol. 14, no. 2, pp. 45–61. DOI: <https://doi.org/10.17759/mda.2024140203> (In Russ., abstr. in Engl.).

References

1. Eileen Nielsen. Practical time series analysis: forecasting with statistics and machine learning.: Per. from English – St. Petersburg.: Dialectics, 2021. – 544 p.
2. Box J., Jenkins G.M. Time series analysis, forecast and management. M.: Mir, 1974. 406 p.
3. J. Vander Plas. Python for Complex Problems: Data Science and Machine Learning. – St. Petersburg: Peter, 2018. – 576 p.
4. Alexey Chernobrovov. Top 7 Python libraries for time series analysis. URL: <https://chernobrovov.ru/articles/top-7-python-bibliotek-dlya-vremennyh-ryadov.htm> l. (Date of reference: 02.03.2024)

***Elena A. Svekolkova**, Student, Department of Mathematical cybernetics, Moscow Aviation Institute (National Research University) (MAI), Moscow, Russia, ORCID: <https://orcid.org/0009-0000-6161-571X>, e-mail: elena.cvekolnikova@gmail.com

****Valentin N. Panovskiy**, Candidate of Physical and Mathematical Sciences, Associate Professor, Department of Mathematical cybernetics, Moscow Aviation Institute (National Research University) (MAI), Moscow, Russia, ORCID: <https://orcid.org/0009-0007-1708-8984>, e-mail: panovskiy.v@yandex.ru



5. S.G. Antipov, M.V. Fomina. The problem of anomaly detection in sets of time series. Program products and systems, No. 2 (98) 2012. – 168 p. – Pp. 78–82.
6. Google Colab: [Electronic resource]. URL: <https://colab.research.google.com/>. (Date of reference: 04.03.2024)
7. Kaggle: [Electronic resource]. URL: <https://www.kaggle.com/>. (Date of reference: 04.03.2024)
8. NumPy: [Electronic resource]. URL: <https://numpy.org/doc/>. (Date of reference: 04.03.2024)
9. Matplotlib: [Electronic resource]. URL: <https://matplotlib.org/stable/index.html>. (Date of reference: 04.03.2024)
10. Seaborn: [Electronic resource]. URL: <https://seaborn.pydata.org/>. (Date of reference: 04.03.2024)
11. Statsmodels: [Electronic resource]. URL: <https://www.statsmodels.org/stable/index.html>. (Date of reference: 04.03.2024)
12. Prophet: [Electronic resource]. URL: <https://prophet.readthedocs.io/en/latest/>. (Date of access: 04.03.2024)
13. Kats: [Electronic resource]. URL: <https://kat.readthedocs.io/en/latest/>. (Date of reference: 04.03.2024)
14. Merlion: [Electronic resource]. URL: <https://opensource.salesforce.com/Merlion/latest/index.html>. (Date of reference: 04.03.2024)
15. DeepAR Forecasting Algorithm: [Electronic resource]. URL: <https://docs.aws.amazon.com/sagemaker/latest/dg/deepar.html>. (Date of access: 04.03.2024)
16. Python: [Electronic resource]. URL: <https://www.python.org/> (Date of access: 04.03.2024)
17. Pismarov A.V. To the question of prediction of the endurance limit of threaded parts with surface hardening // Trudy MAI. 2023. no. 129. URL: <https://trudymai.ru/published.php?ID=173003>
18. Piganov M.N., Kulikov A.V., Novomeisky D.N. Predictional mathematical models of thin-film elements of microassembly // Trudy MAI. 2023. no. 131. URL: <https://trudymai.ru/published.php?ID=175920>
19. Blinov A.V., Razumov D.A. The procedure of formalization of strategies as an element of the methodology for taking into account uncertainty factors in forecasting the indicators of the implementation of space technology development programs // Trudy MAI. 2022. no. 122. URL: <https://trudymai.ru/published.php?ID=164270>
20. Belyaev B.V., Lebedev A.S. Methodology for predicting the residual resource during depressurization of aircraft // Trudy MAI. 2022. no. 125. URL: <https://trudymai.ru/published.php?ID=168167>

Получена 08.04.2024
Принята в печать 23.04.2024

Received 08.04.2024
Accepted 23.04.2024