

УДК 004.9

SARIMA И PROPHET: СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ В ТРАНСПОРТНЫХ СИСТЕМАХ

© Сидорова Е.В., Головнин О.К.

Самарский национальный исследовательский университет
имени академика С.П. Королева, г. Самара, Российская Федерация

e-mail: sidoroeaterina@gmail.com

Высокоточное прогнозирование рисков наступления дорожно-транспортных происшествий (ДТП) является необходимым для разработки новых и совершенствования существующих способов обеспечения безопасности транспортных систем [1]. В данной работе проведено сравнение модели регрессии SARIMA [2] и модели Prophet [3] применительно к задаче прогнозирования ДТП в долгосрочном варианте с последующей проверкой результатов прогнозирования на реальных данных.

Эксперименты проведены на данных из открытого источника Департамента транспорта Великобритании [4], содержащих подробную информацию о ДТП с 2005 по 2019 год. В ходе вычислительного эксперимента прогнозируется количество возможных ДТП моделями SARIMA и Prophet в сравнении с реальными данными.

График зависимости среднего количества ДТП по месяцам от времени приведен на рисунке 1. Количество ДТП с течением времени уменьшается, следовательно, ожидается, что прогноз будет отражать эту тенденцию.

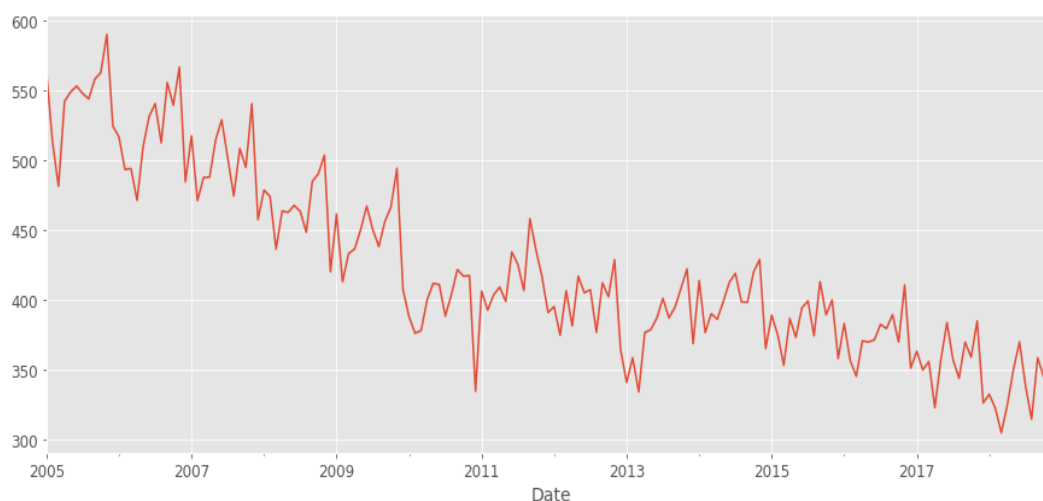


Рис. 1. Зависимость среднего количества ДТП в месяц от времени

В качестве оптимальных параметров будут выступать параметры SARIMA $(0, 1, 1) \times (0, 1, 1, 12)$. Результаты диагностики модели SARIMA отражены на рисунке 2.

На рисунке 2 приведены 4 графика: 1 – график стандартизированных остатков с течением времени; 2 – гистограмма и расчетная плотность стандартизированных остатков, а также нормальная $(0, 1)$ плотность; 3 – нормальный QQ участок с нормальной опорной линией; 4 – коррелограмма. Остатки временного ряда распределены нормально и имеют низкую корреляцию с запаздывающими версиями самих себя. Таким образом, заключим, что модель удовлетворительна.

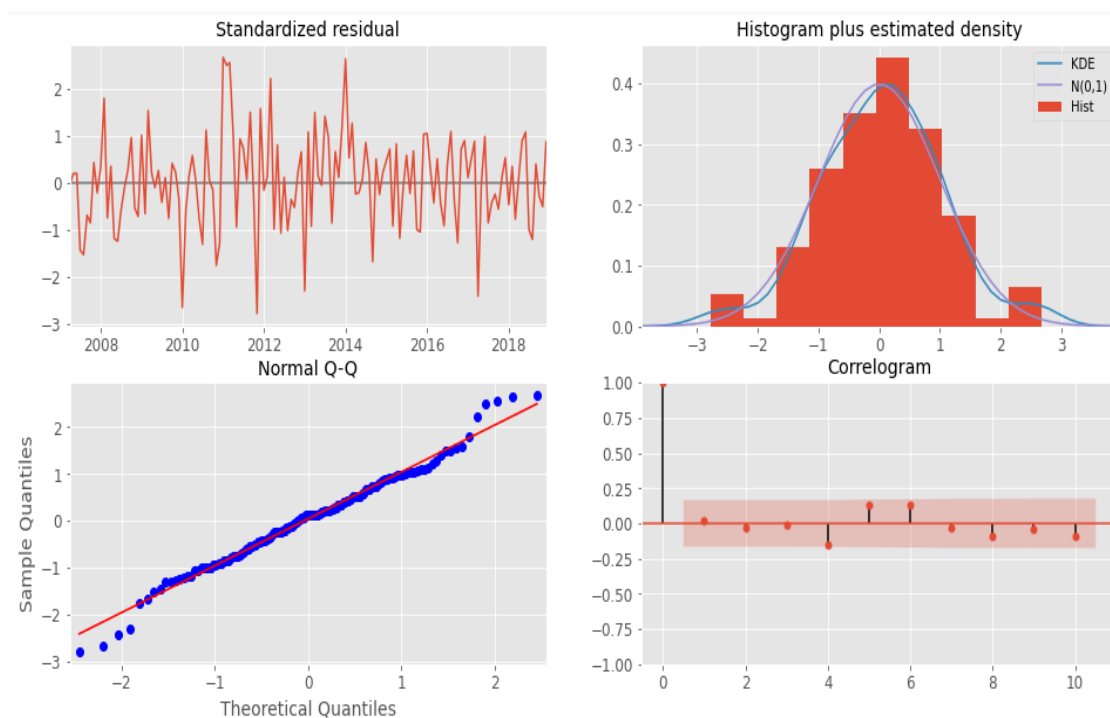


Рис. 2. Результаты диагностики модели SARIMA

Результаты прогнозирования с помощью модели SARIMA на 6 лет начиная с 01.01.2013 приведены на рисунке 3 (синим обозначено прогнозируемое количество ДТП, красным – реальное).

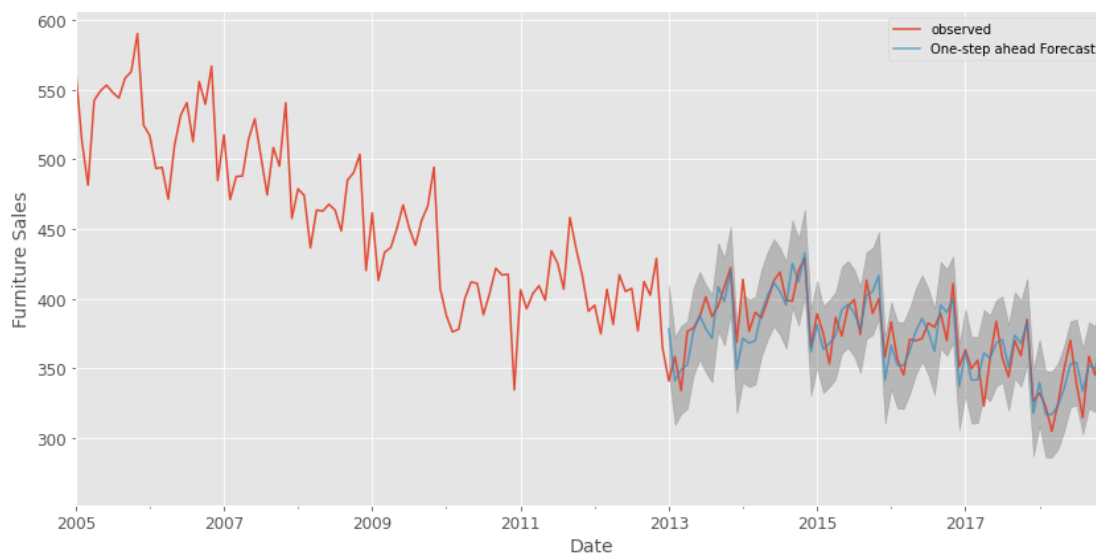


Рис. 3. График зависимости прогнозируемых ДТП от времени

Прогнозы модели отражают общую тенденцию количества ДТП к уменьшению. Среднеквадратичная ошибка $MSE = 209,01$, таким образом, оценщик предсказывает наблюдения параметра с достаточной средней точностью.

Аналогичный вычислительный эксперимент проведен для модели Prophet, которая показала среднеквадратичную ошибку $MSE = 3988,52$.

Таким образом, проведено прогнозирование ДТП с помощью двух моделей. Сравнивая результаты среднеквадратичной ошибки, можно сделать вывод о том, что модель SARIMA показывает более точные результаты по сравнению с моделью Prophet, поэтому дальнейшие исследования будут направлены в сторону совершенствования модели SARIMA.

Библиографический список

1. Головнин О.К. Предсказательная модель выявления возможных угроз безопасности функционирования транспортной системы на основе интеллектуального анализа федеративных данных Smart City // Перспективные информационные технологии: труды междунар. конф. Самара: СНЦ РАН, 2020. С. 248–250.

2. Модель SARIMA и статистика скользящего окна для локальных метеоданных / В.М. Горяев, Д.Б. Бембитов, Д.Н. Мучкаев, В.Х. Аль-Килани // Современные наукоемкие технологии. 2019. № 6. С. 31–38.

3. Прогнозирование временных рядов с помощью Prophet: введение. URL: <https://r-analytics.blogspot.com/2019/08/prophet.html>.

4. Официальный сайт правительства Великобритании. Департамент транспорта. Датасеты. URL: <https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data>.