

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра информатики и программирования

**РАЗРАБОТКА ПРИЛОЖЕНИЯ, РЕАЛИЗУЮЩЕГО
ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ РАЗЛИЧНЫМИ
МЕТОДАМИ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 441 группы

направления 02.03.03 Математическое обеспечение и администрирование
информационных систем

факультета компьютерных наук и информационных технологий

Агапова Максима Леонидовича

Научный руководитель:

доцент, к.ф.-м.н.

К.П. Вахлаева

подпись, дата

Зав. кафедрой:

к.ф.-м.н., доцент

М.В. Огнева

подпись, дата

Саратов 2019

ВВЕДЕНИЕ

Актуальность темы. В бизнес аналитике часто возникают задачи связанные с временными рядами. В качестве примеров таких временных рядов можно привести ряды среднедневных цен на акции определённой компании, среднемесячного уровня безработицы, измеренного в течение нескольких лет, среднегодового уровня производства автомобилей. Интерес представляет задача прогнозирования временных рядов. Подразумевается, что зная значение признака в прошлом, можно предсказать его в будущем. На данный момент существует несколько методов и моделей прогнозирования временных рядов.

Математическая модель – описание какого-либо класса явлений внешнего мира, выраженное с помощью математической символики. Формальная математическая конструкция становится моделью после «наполнения» ее физическим содержанием, указанием связи символов с характеристиками объекта. В данной работе рассматриваются модель Хольта-Винтерса и SARIMA-модель, предназначенные для определения зависимостей величин, характеризующих объект, от времени с целью прогноза их дальнейшего поведения. До 1920-х гг. задача прогноза наблюдаемого процесса решалась методом экстраполяции наблюдаемой временной зависимости, затем появились и получили развитие другие методы, в которых, ограничиваются линейными приближениями [1]-[4]. ARIMA-модели (AutoRegressive Integrated Moving Average) оказались достаточно эффективны для описания различных процессов. Модель Хольта-Винтерса наиболее эффективна в случае краткосрочного прогноза. Они применялись и применяются в технике, метеорологии, экономике, социологии для решения задач прогнозирования и/или управления [5]. В настоящее время многие часто используемые модели прогнозирования временных рядов уже реализованы в языке программирования Python.

Цель бакалаврской работы – изучение и реализация различных формализованных методов и подходов к анализу и прогнозированию временных рядов.

Поставленная цель определила **следующие задачи**:

1. Ознакомиться с различными методами анализа и прогнозирования временных рядов;
2. Построить модель Хольта-Винтерса и SARIMA-модель прогнозирования на основе изученных методов;
3. Подобрать и настроить параметры полученных моделей;
4. Решить задачу прогноза временной последовательности из конкретной предметной области на заданный период с использованием созданных моделей;
5. Провести оценку адекватности прогнозирования.

Методологические основы разработки приложения, реализующего прогнозирование временных рядов различными методами представлены в работах Дж. Бокса, Т. Дженкинса и Д.Н. Хорафаса по анализу и прогнозированию временных рядов [4, 6].

Практическая значимость бакалаврской работы. В ходе выполнения практической части бакалаврской работы были произведены прогнозы временных рядов моделью Хольта-Винтерса и SARIMA-моделью. Также обе модели прошли проверку на адекватность, что доказывает правильность их реализации.

Структура и объём работы. Бакалаврская работа состоит из введения, 10 разделов, заключения, списка использованных источников и приложения. Общий объем работы – 62 страницы, из них 47 страниц – основное содержание, включая 28 рисунков и 1 таблицу, список использованных источников информации – 22 наименования.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Временные ряды» посвящён описанию временных рядов и его основных компонентов, классификация методов и моделей прогнозирования [6]-[8].

Второй раздел «Метод скользящего среднего» посвящен описанию метода скользящего среднего порядка q . В ней предполагается, что значение ряда y_t – это линейная комбинация q последних значений шумовой компоненты. Также была продемонстрирована возможность сглаживания исходного ряда для выявления трендов с помощью данного метода.

Третий раздел «Метод экспоненциального сглаживания» посвящен описанию метода экспоненциального сглаживания для выравнивания ряда. А также изучен метод двойного экспоненциального сглаживания, использующийся при наличии тренда [9].

Четвертый раздел «Тройное экспоненциальное сглаживание» посвящен описанию модели Хольта-Винтерса. Идея этого метода заключается в добавлении еще одной, третьей, компоненты - сезонности. Сезонная компонента в модели будет объяснять повторяющиеся колебания вокруг уровня и тренда, а характеризоваться она будет длиной сезона - периодом, после которого начинаются повторения колебаний [10].

Пятый раздел «Оценка качества прогнозируемых значений с использованием кросс-валидации» посвящен методу кросс-валидации для поиска и оценки качества параметров α , β , γ [11].

Шестой раздел «SARIMA-модель» посвящен описанию SARIMA-модели. Моделью SARIMA порядка $(p,d,q) \times (P,D,Q)$ называется модель SARMA порядка $(p,q) \times (P,Q)$ для ряда, к которому d раз было применено обычное дифференцирование и D раз сезонное [12]-[14].

Седьмой раздел «Прогнозирование SARIMA-модели» посвящен описанию последовательных действий для осуществления прогнозирования временного ряда с помощью SARIMA-модели.

Восьмой раздел «Проверка на стационарность и реализация SARIMA-модели» посвящен проверке временного ряда на стационарность с помощью критерия Дики-Фуллера для поиска единичных корней [15].

Девятый раздел «Оценка адекватности прогнозирования. Анализ остатков» посвящен оценке адекватности прогнозирования и анализу остатков. Анализ остатков - это техника, которая помогает понять, есть ли у прогнозирующей модели небольшие недостатки, которые можно устранить доработкой, или же фундаментальные проблемы. Остатки – это разность между фактом и прогнозом. Остатки оценивают ошибку, то есть шумовую компоненту.

Десятый раздел «Построение прогноза временного ряда с помощью модели Хольта-Винтерса и SARIMA-модели» посвящен описанию реализации прогнозирования временных рядов с помощью модели Хольта-Винтерса и SARIMA-модели.

Подраздел «Формальная постановка задачи» описывает общую задачу прогнозирования. В качестве примера временного ряда были рассмотрены данные о количестве просмотров рекламы и число покупок игровой валюты в приложении «Clash of Clans», взятые с портала www.appannie.com [16].

Подраздел «Сглаживание трендов с помощью метода скользящего среднего» показывает результат сглаживания тренда при выборе разной ширины интервала. В библиотеке Pandas имеется готовая реализация метода скользящего среднего - `DataFrame.rolling(window).mean()` [17].

Подраздел «Модель Хольта-Винтерса» показывает практическую реализацию данной модели. Продемонстрирован прогноз для двух временных рядов, поиск наиболее подходящих параметров с помощью кросс-валидации, а также выявлены имеющиеся аномалии [18]-[20].

Подраздел «Реализация SARIMA-модели» показывает практическую реализацию SARIMA -модели. С помощью теста Дики-Фуллера, функции автокорреляции и частичной автокорреляции был проанализирован временной ряд количества просмотров рекламы. Методом Акаике были подобраны параметры $(p,d,q) \times (P,D,Q)$. А получившийся прогноз показал средний процент отклонения ошибок равный 3.94% [21].

Подраздел «Оценка адекватности прогнозирования» описывает получившиеся графики прогнозируемого отклонения временного ряда просмотров рекламы и покупок игровой валюты для модели Хольта-Винтерса. Для проверки адекватности Sarima-модели надо сначала получить остатки модели и построить для них ACF.

ЗАКЛЮЧЕНИЕ

В работе были изучены подходы к использованию дискретных последовательностей экспериментальных данных (временных рядов) для конструирования моделей Хольта-Винтерса и SARIMA-модели, предназначенных для прогноза поведения объекта. Анализ среднего процента отклонения ошибок прогноза SARIMA-модели показал лучше результат чем у модели Хольта-Винтерса. В то же время реализация ARIMA-модели оказалась более трудоемкой. Для реализации выбранных моделей использовалась командная оболочка для интерактивных вычислений Jupyter Notebook, а языком программирования был выбран Python. Основной библиотекой для работы с временными рядами являлись Statsmodels. Использование двух временных рядов показало зависимость модели Хольта-Винтерса от объема прогнозируемых данных. Прогноз данных о количестве просмотров рекламы показал средний процент отклонения ошибок 4.84%, а прогноз числа покупок игровой валюты представил средний процент отклонения ошибок равный 16.88%.

Основные источники информации:

1. Структурные средние величины. [Электронный ресурс]. URL: <http://www.grandars.ru/student/statistika/strukturnye-srednie-elichiny.html> (Дата обращения 20.05.2019).
2. Оценка точности прогнозирования случайной величины. [Электронный ресурс]. URL: <https://www.economy-web.org/?p=289> (Дата обращения 20.05.2019).
3. Significance of ACF and PACF Plots In Time Series Analysis. [Электронный ресурс]. URL: <https://towardsdatascience.com/significance-of-acf-and-pacf-plots-in-time-series-analysis-2fa11a5d10a8> (Дата обращения 20.05.2019).
4. Бокс Дж., Дженкинс Т. Анализ временных рядов. Прогноз и управление. М.: Мир, 1974. (Дата обращения 15.10.2018).
5. Методы анализа временных рядов: сглаживание. [Электронный ресурс]. URL: <http://www.prognoz.ru/blog/platform/time-series-manual-1/> (Дата обращения 20.05.2019).
6. Хорафас Д.Н. Системы и моделирование. М.: Мир, 1967. (Дата обращения 15.10.2018).
7. Модель прогнозирования временных рядов по выборке максимального подобия. [Электронный ресурс]. URL: http://www.mbureau.ru/articles/dissertaciya-model-prognozirovaniya-vremennyh-ryadov-glava-1#p_1.3 (Дата обращения 20.05.2019).
8. Классификация методов и моделей прогнозирования. [Электронный ресурс]. URL: <https://habr.com/ru/post/177633/> (Дата обращения 20.05.2019).
9. Модель экспоненциального сглаживания. [Электронный ресурс]. URL: <https://4analytics.ru/prognozirovanie/malo-dannix-dlya-progoza-model-eksponencialnogo-sglajivaniya.html> (Дата обращения 20.05.2019).
10. Прогноз по методу экспоненциального сглаживания с трендом и сезонностью Хольта-Винтерса. [Электронный ресурс]. URL:

<https://4analytics.ru/prognozirovanie/prognoz-po-metodu-eksponencialnogo-sglajvaniya-s-trendom-i-sezonnostyu-xolta-vintersa.html> (Дата обращения 20.05.2019).

11. Кросс-валидация. [Электронный ресурс]. URL: <http://long-short.pro/post/kross-validatsiya-cross-validation-304/> (Дата обращения 20.05.2019).

12. Прогнозирование временных рядов. [Электронный ресурс]. URL: http://math.spbu.ru/SD_AIS/documents/2014-12-341/2014-12-tw-13.pdf (Дата обращения 20.05.2019).

13. Модель ARMA. [Электронный ресурс]. URL: https://studme.org/72683/ekonomika/model_arma (Дата обращения 20.05.2019).

14. Машинное обучение и анализ данных. [Электронный ресурс]. URL: <https://www.coursera.org/lecture/data-analysis-applications/vriemiennyye-riady-EjNEV> (Дата обращения 15.10.2018).

15. Критерий Дикки-Фуллера проверки наличия единичных корней. [Электронный ресурс]. URL: <http://be5.biz/ekonomika/e008/85.html> (Дата обращения 20.05.2019).

16. Clash of Clans. Рейтинг приложения. [Электронный ресурс]. URL: <https://www.appannie.com/apps/ios/app/clash-of-clans/rank-history/?vtype=day&countries=US&device=iphone&view=grossing&legends=2222&date=2019-04-21~2019-05-20> (Дата обращения 20.05.2019).

17. Проблема обнаружения аномалий в наборах временных рядов. [Электронный ресурс]. URL: <http://www.swsys.ru/index.php?page=article&id=3117> (Дата обращения 20.05.2019).

18. Алгоритм краткосрочного прогнозирования временных рядов. [Электронный ресурс]. URL: <http://elib.spbstu.ru/dl/2/v17-3736.pdf/download/v17-3736.pdf> (Дата обращения 20.05.2019).

