|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

**КУРСОВАЯ РАБОТА**

по дисциплине: «Языки программирования для статистической   
обработки данных»

**Тема курсовой работы**: «Разработка программы для прогнозирования временного ряда на основе линейных авторегрессионных моделей на основе данных статусов авиаперевозок»

|  |  |  |
| --- | --- | --- |
| Студент группы ИМБО-02-22 | Ким Кирилл Сергеевич | (подпись) |
| Руководитель  курсовой работы | старший преподаватель,  Юрченков И.А. | (подпись) |
| Работа представлена к защите | «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_2024 г. |  |
| Допущен к защите | «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_2024 г. |  |

Москва 2024 г.

|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

|  |  |  |
| --- | --- | --- |
|  |  | Утверждаю |
|  |  | И.о. заведующего кафедрой ПМ |
|  |  | \_\_\_\_\_\_\_\_\_\_\_\_\_\_Смоленцева Т.Е.  (подпись) |
|  |  | «09» февраля 2024 г. |

**ЗАДАНИЕ**

**на выполнение курсовой работы**

по дисциплине «Языки программирования для статистической обработки данных»

|  |  |  |  |
| --- | --- | --- | --- |
| Студент Ким Кирилл Сергеевич | | | Группа ИМБО-02-22 |
| **Тема** «Разработка программы для прогнозирования временного ряда на основе линейных авторегрессионных моделей на основе данных статусов авиаперевозок» | | | |
| **Исходные данные:** выбранная студентом задача и алгоритм её решения, а также набор данных | | | |
| **Перечень вопросов, подлежащих разработке, и обязательного графического материала:**  Описание решаемой задачи машинного обучения или статистической обработки данных (математическая формулировка, проблематика, существующие способы решения)  Анализ выбранного алгоритма или метода решения выбранной задачи  Выбор данных и описание набора данных, описание предикторов и целевых факторов  Построение сценария и логики обработки данных на основе выбранного алгоритма или метода Оценка качества решения задачи на основе метрик качества | | | |
| **Срок представления к защите курсовой работы: до** «24» мая 2024 г. | | | |
| **Задание на курсовую работу выдал** | *Подпись руководителя* | Юрченков И.А.  *(ФИО руководителя)* | |
| «09» февраля 2024 г. | | | |
| **Задание на курсовую работу получил** | *Подпись обучающегося* | Ким К.С.  *(ФИО обучающегося)* | |
| «09» февраля 2024 г. | | | |

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 4](#_Toc165571391)

[1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ 5](#_Toc165571392)

[1.1 Основные понятия временных рядов 5](#_Toc165571393)

[1.2 Линейная авторегрессионная модель 7](#_Toc165571394)

[2 ПРАКТИЧЕСКАЯ ЧАСТЬ 10](#_Toc165571395)

[2.1 Линейная авторегрессионная модель на основе данных авиаперевозок для прогнозирования временного ряда 10](#_Toc165571396)

[ЗАКЛЮЧЕНИЕ 11](#_Toc165571397)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 12](#_Toc165571398)

[Теоретическая часть 12](#_Toc165571399)

[Практическая часть 12](#_Toc165571400)

[ПРИЛОЖЕНИЯ 13](#_Toc165571401)

[Листинг 13](#_Toc165571402)

# ВВЕДЕНИЕ

В современном мире авиационная отрасль играет ключевую роль в международных и национальных перевозках. Эффективное управление авиаперевозками требует точного прогнозирования временных рядов статусов авиаперевозок, таких как задержки, отмены рейсов и другие изменения. Разработка программы для прогнозирования временного ряда на основе линейных авторегрессионных моделей позволит улучшить планирование и управление авиаперевозками.

Целью данного исследования является разработка программы для прогнозирования временного ряда на основе линейных авторегрессионных моделей на основе данных статусов авиаперевозок. Для достижения этой цели поставлены следующие задачи:

1. Изучить теоретические основы временных рядов и авторегрессионных моделей.
2. Провести анализ данных статусов авиаперевозок.
3. Разработать программу прогнозирования временного ряда на основе полученных данных.

После завершения данного исследования будет создана программа, способная предсказывать статусы авиаперевозок с высокой точностью, что поможет авиакомпаниям оптимизировать свою деятельность и повысить уровень обслуживания пассажиров.

# 1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

## 1.1 Основные понятия временных рядов

Для разработки программы прогнозирования временного ряда на основе данных статусов авиаперевозок необходимо разобраться с основными понятиями временных рядов. Временной ряд — это последовательность данных , измеренных в последовательные моменты времени . Он может быть стационарным или нестационарным, что необходимо учитывать при выборе модели.

Временной ряд называется стационарным

* в узком смысле, если для любых вектор совпадает по распределению с , то есть при сдвиге всех моментов времени на одно и тоже число совместное распределение значений временного ряда в эти моменты времени не поменяется;
* в широком смысле, если
  + для любого ;
  + не зависит от , то есть в среднем значение временного ряда постоянно;
  + для любых , то есть значение автокорреляции зависит только от длины отрезка времени между двумя значениями;
* для гауссовских распределений, то есть для случая, когда все векторы вида имеют нормальное распределение, определения эквивалентны. Это следует из того, что распределение гауссовского случайного вектора полностью определяется математическим ожиданием и ковариациями;

В нестационарных временных рядах статистические свойства меняются со временем. Они показывают сезонные эффекты, тренды и другие структуры, которые зависят от временного показателя.

Каждое измерение представляет собой наблюдение для конкретной переменной в определенный момент времени.

Компоненты временного ряда:

1. Тренд (*T* — trend) — долгосрочное изменение уровня значений временного ряда. Тренд может быть восходящим, нисходящим или стационарным.
2. Сезонность (*S* — seasonal) — циклическое повторение паттернов в данных с постоянным интервалом времени. Например, продажи игрушек могут иметь сезонное изменение в преддверии праздников.
3. Цикл (*C* — cyclic) — периодические колебания в данных, обычно с более длительным циклом, чем у сезонности. Например, экономические циклы имеют периодичность в несколько лет.
4. Шум (*E* — errors) — непредсказуемая случайная переменная, которая не может быть объяснена трендом, сезонностью или циклом. Шум включает в себя случайные флуктуации и ошибки измерения.

Включение компонентов в модель временного ряда:

1. Аддитивная модель: Аддитивная модель используется, если амплитуда колебаний более-менее постоянная.
2. Мультипликативная модель: Мультипликативная — если амплитуда колебаний зависит от значения сезонной компоненты.

Выбор между аддитивной и мультипликативной моделями зависит от того, как компоненты взаимодействуют друг с другом в конкретном временном ряде.

Анализ временных рядов позволяет выявлять закономерности в данных, делать прогнозы и принимать более обоснованные решения на основе исторических данных. Методы анализа временных рядов включают в себя статистические и математические модели, такие как ARIMA (среднее, интегрированное, скользящее среднее), экспоненциальное сглаживание и регрессионный анализ.

## 1.2 Линейная авторегрессионная модель

Авторегрессия (autoregressive model, AR) — это регрессия ряда на собственные значения в прошлом. Другими словами, наши признаки в модели обычной регрессии мы заменяем значениями той же переменной, но за предыдущие периоды.

Когда мы прогнозируем значение в период *t* с помощью данных за предыдущий период (AR(1)), уравнение будет выглядеть следующим образом.

где — это константа, — вес модели, — значение в период

Количество используемых предыдущих периодов определяется параметром *p*. Обычно записывается как AR(*p*).

Модель скользящего среднего (moving average, MA) помогает учесть случайные колебания или отклонения (ошибки) истинного значения от прогнозного. Можно также сказать, что модель скользящего среднего — это авторегрессия на ошибку.

Если использовать ошибку только предыдущего наблюдения, то уравнение будет выглядеть следующим образом.

где — это среднее значение временного ряда, — вес модели, — ошибка в период .

Такую модель принято называть моделью скользящего среднего с параметром *q* = 1 или MA(1). Разумеется, параметр q может принимать и другие значения (MA(*q*)).

ARMA предполагает, что в данных отсутствует тренд и сезонность (данные стационарны). Если данные нестационарны, нужно использовать более сложные версии этих моделей:

Модель ARMA(*p, q*) по сути является суммой моделей AR(*p*) и MA(*q*), иначе говоря, модель есть сумма нескольких предыдущих значений ряда и нескольких предыдущих значений белого шума с некоторым коэффициентами.

Эквивалентную запись ряда в терминах оператора сдвига можно получить, рассмотрев два многочлена

или

где , и . Заметим, что во втором представлении константа заменена на . На самом деле, стационарность такого ряда будет определяться только его AR(*p*) компонентой, то есть значениями коэффициентов, так ряд в модели MA(*q*) всегда является стационарным.

* ARIMA, здесь добавляется компонент Integrated (I), который отвечает за удаление тренда (сам процесс называется дифференцированием); и
* SARIMA, эта модель учитывает сезонность (Seasonality, S)
* SARIMAX включает еще и внешние или экзогенные факторы (eXogenous factors, отсюда и буква X в названии), которые напрямую не учитываются моделью, но влияют на нее.

Параметров у модели SARIMAX больше. Их полная версия выглядит как SARIMAX(p, d, q) x (P, D, Q, s). В данном случае, помимо известных параметров p и q, у нас появляется параметр d, отвечающий за тренд, а также набор параметров (P, D, Q, s), отвечающих за сезонность.

Теперь давайте воспользуемся моделью SARIMAX для прогнозирования авиаперевозок.

# 2 ПРАКТИЧЕСКАЯ ЧАСТЬ

## Линейная авторегрессионная модель на основе данных авиаперевозок для прогнозирования временного ряда

# ЗАКЛЮЧЕНИЕ

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

## Теоретическая часть

* 1. Волков Никита. Аналитика временных рядов / Учебник по машинному обучению [Электронный ресурс]. https://education.yandex.ru/handbook/ml/article/analitika-vremennyh-ryadov
  2. Волков Никита. Аналитика временных рядов / Учебник по машинному обучению [Электронный ресурс]. https://education.yandex.ru/handbook/ml/article/modeli-vida-arima

## Практическая часть

# ПРИЛОЖЕНИЯ

## Листинг