

Лабораторная работа № 2
Прогнозирование коротких временных рядов
Вариант № 4

Цель работы

Произвести линейный регрессионный анализ временного ряда.

Задание

Исходные данные - временной ряд месячного производства молока за 14 лет. Необходимо построить прогноз на 8 месяцев.

Решение

Для решения задачи были использованы библиотеки:

- Matplotlib - библиотека для графического представления данных.
- Pandas - программная библиотека на языке Python для обработки и анализа данных. Она представляет собой специальные структуры данных и операции для манипулирования числовыми таблицами и временными рядами.
- Statsmodels – библиотека на Python, которая позволяет оценивать и анализировать различные статистические модели. Он построен на числовых и научных библиотеках, таких как NumPy и SciPy.

Построим график исходного временного ряда месячного производства молока, пользуясь предоставленными исходными данными.

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import matplotlib
4 matplotlib.use('TkAgg')
5
6 data = pd.read_csv('C:/Users/solidus66/OneDrive/BFY/4 курс 1 сем/ТМ/лаб2/milk.csv')
7
8 plt.figure(figsize=(10, 6))
9 plt.plot(data['milk'], marker='o')
10 plt.title('Original Time Series of Monthly Milk Production')
11 plt.xlabel('Month')
12 plt.ylabel('Milk Production')
13 plt.show()
```

Рисунок 1 – Фрагмент кода. Чтение файла и отрисовка графика
исходного временного ряда месячного производства молока

Запускаем написанный код и получаем график:

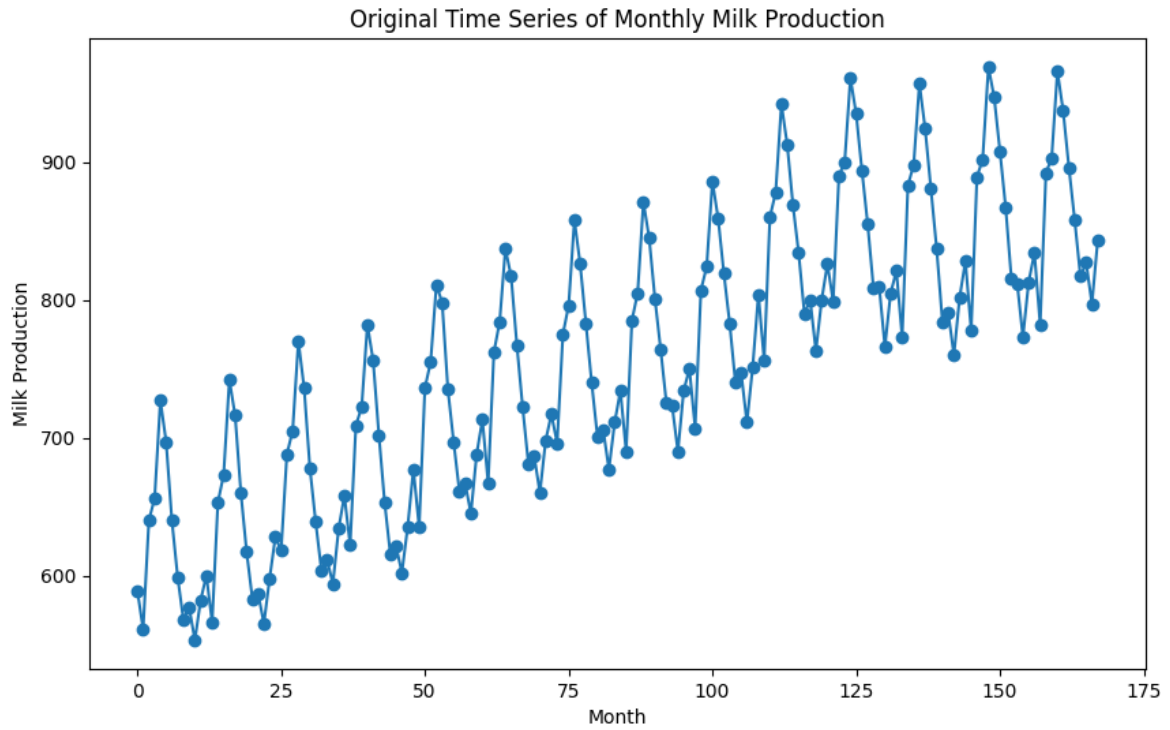


Рисунок 2 – График исходного временного ряда месячного
производства молока

Из графика видим:

Тренд: в целом производство молока имеет тенденцию к увеличению со временем.

Сезонность: есть некоторая сезонность.

Изменение характера: на первый взгляд нет явных резких изменений характера.

Выбросы: сильно выбивающихся из ряда значений не имеется.

После этого анализа можно приступать к прогнозу. Перепишем код. Будем использовать SARIMA (Seasonal AutoRegressive Integrated Moving Average) модель, учитывающую сезонность.

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 from statsmodels.tsa.statespace.sarimax import SARIMAX
4 import matplotlib
5 matplotlib.use('TkAgg')
6
7 data = pd.read_csv('C:/Users/solidus66/OneDrive/BГV/4 курс 1 сем/ТММ/lab2/milk.csv')
8
9 model = SARIMAX(data['milk'], order=(1, 1, 1), seasonal_order=(1, 1, 1, 12))
10 results = model.fit(dispatch=False)
11
12 forecast_steps = 8
13 forecast = results.get_forecast(steps=forecast_steps)
14
15 plt.figure(figsize=(12, 6))
16 plt.plot(data['milk'], label='Initial data', marker='o')
17 plt.plot(forecast.predicted_mean, label='Forecast', color='red', linestyle='dashed', marker='o')
18 plt.title('Monthly milk production forecast')
19 plt.xlabel('Month')
20 plt.ylabel('Milk Production')
21 plt.legend()
22 plt.show()
23
24 forecast_values = forecast.predicted_mean[-forecast_steps:]
25 forecast_values_with_index = forecast_values.reset_index()
26 forecast_values_with_index.columns = ['Month', 'Production forecast']
27 forecast_values_with_index['Month'] += 1
28
29 print("Forecast for 8 months ahead:")
30 print(forecast_values_with_index)
```

Рисунок 3 – Фрагмент кода. Новый код, в котором уже происходит прогноз

В результате работы нового кода, получаем новый график:

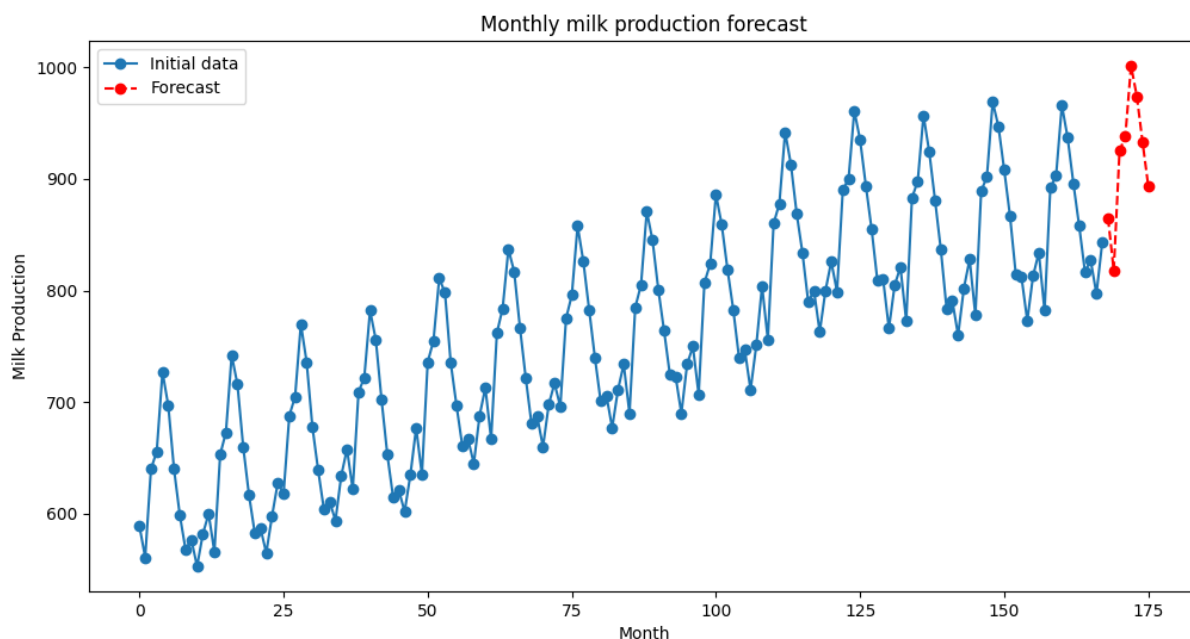


Рисунок 4 – График прогноза месячного производства молока

Синим на графике отображены исходные данные, красным – прогноз на следующие 8 месяцев. Более детальная информация выводится в консоль:

```
Forecast for 8 months ahead:
```

| | Month | Production forecast |
|---|-------|---------------------|
| 0 | 169 | 865.174111 |
| 1 | 170 | 818.032984 |
| 2 | 171 | 925.009744 |
| 3 | 172 | 938.050293 |
| 4 | 173 | 1001.299339 |
| 5 | 174 | 974.036500 |
| 6 | 175 | 932.842537 |
| 7 | 176 | 893.109038 |

Рисунок 5 – Прогноз месячного производства молока в числах