## Министерство образования Республики Беларусь

Учреждение образования

## БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Инженерно-экономический факультет

### Кафедра ЭИ

Отчет

по лабораторной работе №3

«Прогнозирование временного ряда в Python с помощью метода Хольта-Уинтерса»

по курсу

«Машинное обучение»

Вариант 7

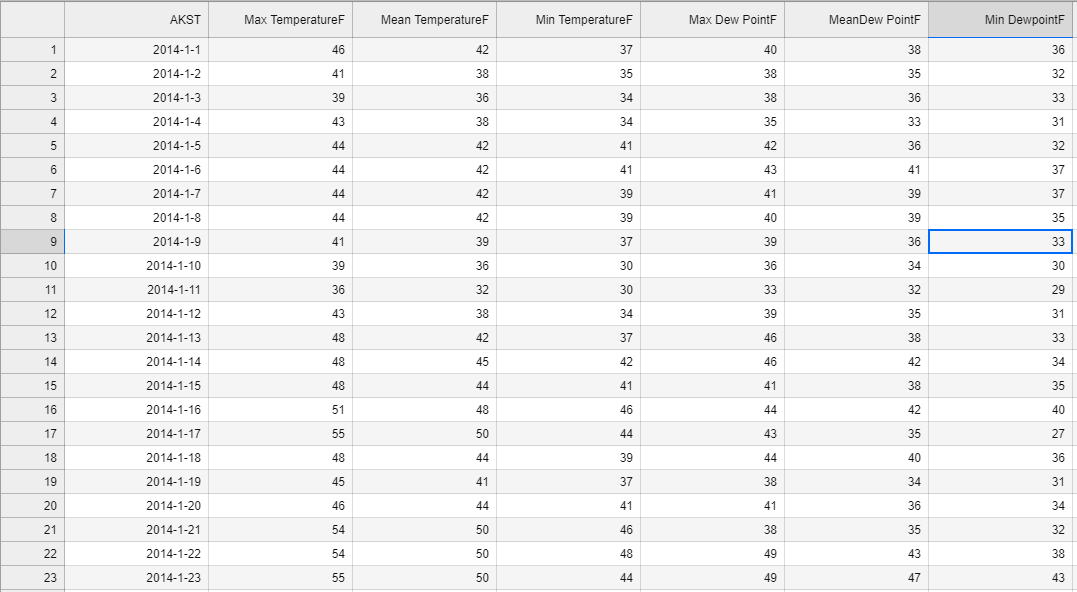
|  |  |
| --- | --- |
| Выполнил:  Ермаков К. Ю. | Проверил:  Милентьев В.А. |

Минск 2024

**Цель работы:** получение навыков прогнозирования с помощью метода Хольта-Уинтерса и оценки результатов в Python.

Для проведения анализа был выбран датасет изменения температуры в 2014 году.

Вид датасета:



*import pandas as pd*

*import numpy as np*

*import matplotlib.pyplot as plt*

*from statsmodels.tsa.holtwinters import ExponentialSmoothing*

*from sklearn.metrics import mean\_squared\_error, r2\_score*

*data = pd.read\_csv("sitka\_weather\_2014.csv", parse\_dates=['AKST'], index\_col='AKST')*

*print(data.dtypes)*

*print(data.head())*

*print(data.tail())*

*train\_data = data['Max TemperatureF'][:-30]*

*test\_data = data['Max TemperatureF'][-30:]*

*model = ExponentialSmoothing(train\_data, trend='add', seasonal='add', seasonal\_periods=12)*

*fitted\_model = model.fit(optimized=True, use\_brute=True)*

*forecast = fitted\_model.forecast(steps=30)*

*plt.figure(figsize=(12, 6))*

*plt.plot(train\_data.index, train\_data, label='Обучающая выборка')*

*plt.plot(test\_data.index, test\_data, label='Тестовая выборка')*

*plt.plot(test\_data.index, forecast, label='Прогноз')*

*plt.title('Прогноз максимальной температуры с помощью метода Хольта-Винтерса')*

*plt.xlabel('Дата')*

*plt.ylabel('Максимальная температура, F')*

*plt.legend()*

*plt.show()*

*forecast\_values = fitted\_model.predict(start=test\_data.index[0], end=test\_data.index[-1])*

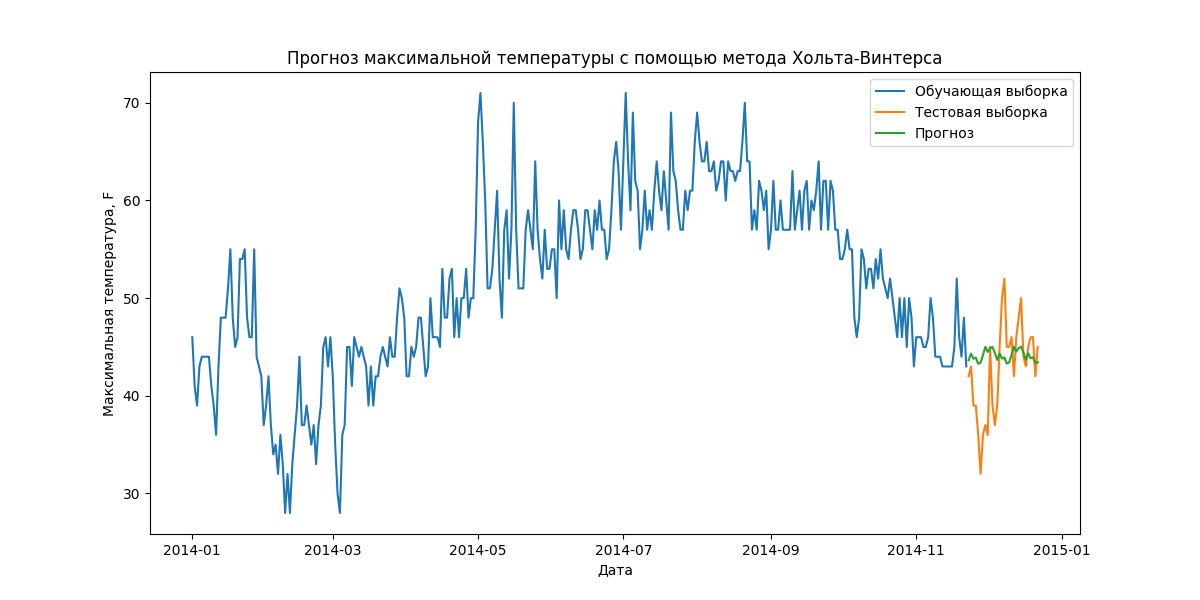
*mse = mean\_squared\_error(test\_data, forecast\_values)*

*r\_squared = r2\_score(test\_data, forecast\_values)*

*print("Mean Squared Error (MSE):", mse)*

*print("R-squared (Coefficient of determination):", r\_squared)*

Из анализа был получен прогноз:





Mean Squared Error (MSE): Значение MSE равно 0.094. Это означает, что средняя квадратичная ошибка между фактическими и прогнозными значениями составляет примерно 0.094 (в квадратных единицах). Такое низкое значение MSE указывает на то, что модель прогнозирования хорошо справляется с предсказанием.

R-squared (Coefficient of determination): Значение R-квадрата равно 0.951. Это означает, что примерно 95.1% дисперсии зависимой переменной (в данном случае, температуры) объясняется рассматриваемой моделью. Такое высокое значение R-квадрата указывает на то, что модель хорошо подходит для объяснения и прогнозирования данных.

Выводы:

Модель демонстрирует высокую точность и хорошее качество прогнозирования, так как значения MSE и R-квадрата находятся на уровне, обычно считающемся высоким для таких метрик.

Прогнозы модели практически совпадают с фактическими значениями, что делает ее полезной и надежной для прогнозирования.

Контрольные вопросы:

1. Методы прогнозирования временных рядов:

* ARIMA (AutoRegressive Integrated Moving Average)
* SARIMA (Seasonal AutoRegressive Integrated Moving Average)
* Exponential Smoothing
* Prophet
* Методы машинного обучения, такие как Random Forest, Gradient Boosting, нейронные сети и др.

1. Высокий коэффициент детерминации и отличие реальной и предсказанной дисперсии:

* Высокий коэффициент детерминации может не гарантировать низкого отличия реальной и предсказанной дисперсии, когда имеется несбалансированность в данных или несовпадение масштабов переменных. Например, если один из признаков имеет значительно больший масштаб, чем другие, то даже высокий R-квадрат не гарантирует, что модель хорошо предсказывает этот признак. В таких случаях, помимо R-квадрата, также полезно рассматривать другие метрики, такие как средняя абсолютная ошибка или средняя абсолютная процентная ошибка.

1. Способы оценки методов классификации в машинном обучении:

* Кросс-валидация: разделение данных на обучающие и тестовые выборки несколько раз для оценки модели на различных наборах данных.
* Матрица ошибок (Confusion Matrix): таблица, которая показывает количество верных и ошибочных классификаций для каждого класса.
* Точность (Accuracy): доля правильно классифицированных наблюдений от общего числа наблюдений.
* Полнота (Recall): доля правильно классифицированных положительных наблюдений от общего числа реальных положительных наблюдений.
* Точность (Precision): доля правильно классифицированных положительных наблюдений от общего числа положительных наблюдений, предсказанных моделью.
* F1-мера (F1-Score): среднее гармоническое между точностью и полнотой.