**Міністерство освіти і науки України**

**Національний технічний університет України «КПІ» імені Ігоря Сікорського**

**Кафедра обчислювальної техніки ФІОТ**

**ЗВІТ**

**з лабораторної роботи №6**

**з навчальної дисципліни «Вступ до технології Data Science»**

**Тема:**

**ДОСЛІДЖЕННЯ ПРОЦЕСІВ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ЗАДАЧ АНАЛІЗУ ДАНИХ**

**Виконав:**

Студент X курсу кафедри ОТ ФІОТ,

Навчальної групи ІА-73

Петренко П.П.

**Перевірив:**

Професор кафедри ОТ ФІОТ

Писарчук О.О.

**Київ 2023**

**І. Мета:**

виявити дослідити та узагальнити особливості застосування штучного інтелекту для аналізу даних з використанням спеціалізованих пакетів мови програмування Python.

**ІІ. Завдання:**

Лабораторія провідної ІТ-компанії реалізує проект із розробки системи підтримки прийняття рішень для автоматизації процесів вироблення рішень у сфері задач електронної комерції.

Замовниками програмної системи – низка оптових компаній.

Вам, як Data Science [Engineer](https://jobs.dou.ua/companies/gravitum/vacancies/147764/) поставлене наступне завдання.

Розробити скрипт мовою Python що реалізує прогнозування динаміки зміни контрольованого процесу відповідно до технічних умов, заданих у таблиць Д1, Д2 додатку 1.

Завдання реалізувати у відповідності до пунктів:

* 1. Здійснити прогнозування динаміки зміни досліджуваного процесу;
  2. Здійснити розрахунок статистичних характеристик результатів прогнозування;
  3. Порівняти статистичні характеристики та прогнозовані оцінки з фактичними даними (модельними, або тими, що отримані в процесі спостереження) та зробити висновки.

Результат представити у формі:

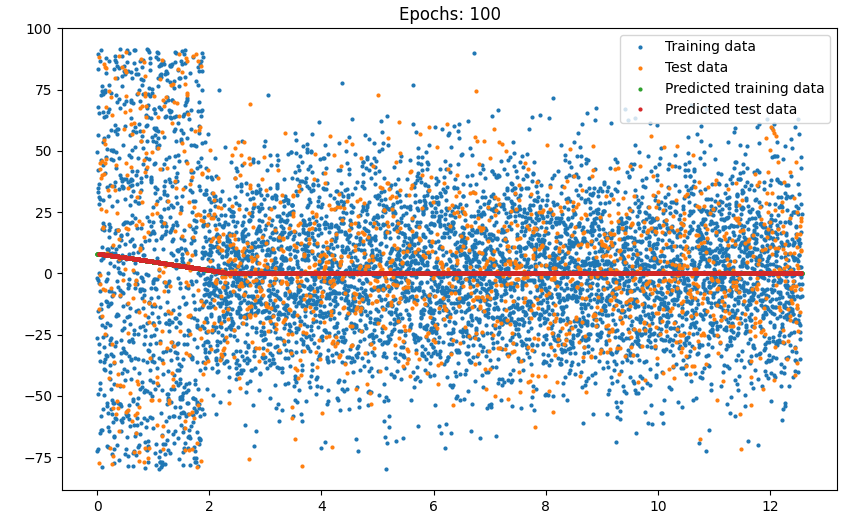
1.5. Результати архітектурного проектування скрипта, що реалізує технічні умови задачі.

1.6. Програмний скрипт, результати його функцілнування.

1.7. Результати дослідження ефективності та їх аналіз.

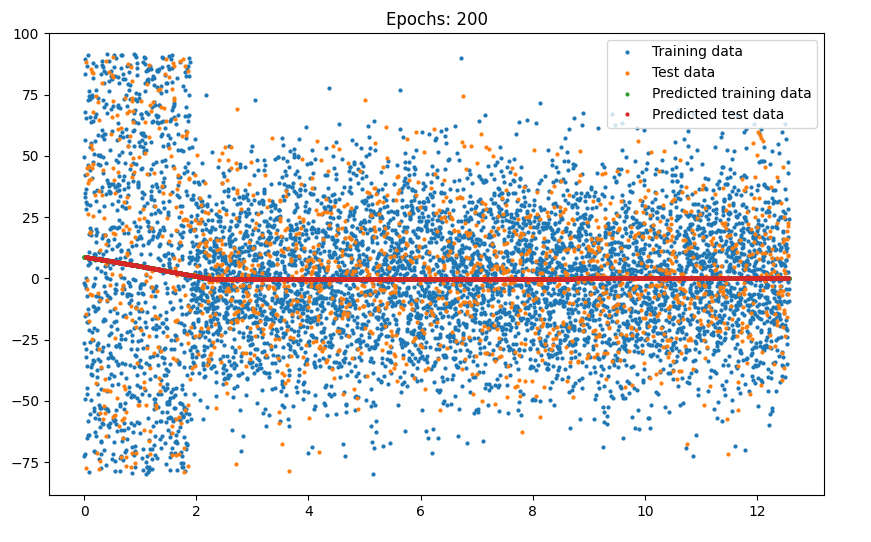
| 6, 21 | Розробити програмний скрипт, що реалізує:  1. Модель зміни досліджуваного процесу – dataset за законом косинуса, у 8000 дискретних вимірів, з нормальним шумом (нульове середнє та СКВ похибки 24) та 15% аномальних вимірів з рівномірним розподілом в межах вибірки;  2. Штучну нейронну мережу з прогнозування даних відповідно до параметрів заданого в п.1 dataset;  3. Відображення процесу прогнозування у формі графіків.  4. Дослідити залежність точності прогнозування числового ряду від кількості епох навчання. |
| --- | --- |

| **6, 21** | **Розробити програмний скрипт, що реалізує:**  **1. Отримання даних коливання обраної валюти з ресурсів Інтернет за обраний проміжок часу;**  **2. Прогнозування зміни курсу обраної валюти на 3 інтервали спостереження з використанням обраного статистичного методу прогнозування.**  **3. Прогнозування зміни курсу обраної валюти на 3 інтервали спостереження з використанням штучних нейронних мереж.**  **4. Порівняння та відображення статистичних характеристик точності і абсолютних значень прогнозованих оцінок на останній точки часового інтервалу прогнозу.**  **Сформувати обґрунтовані висновки про вибір доцільного методу прогнозування.** |
| --- | --- |

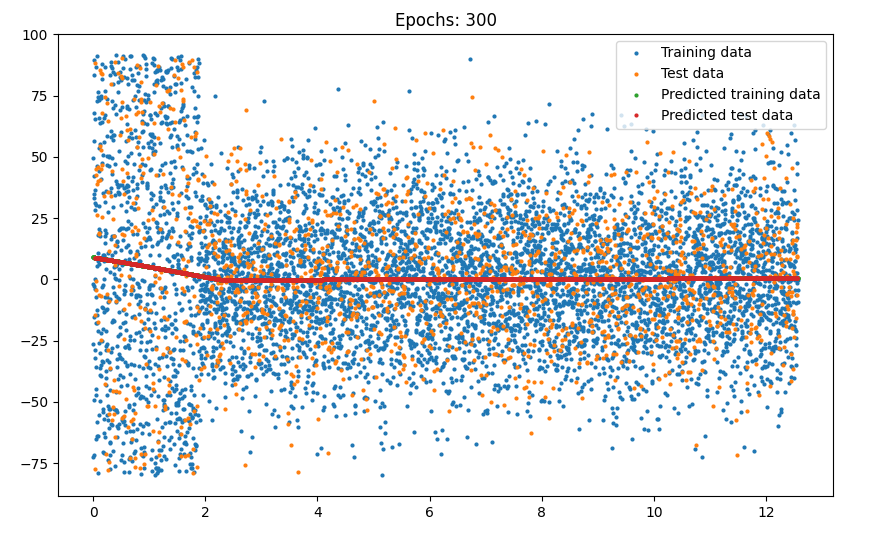
****

****

**Рисунок 1 мережа MLP тренована упродовж 100 епох**

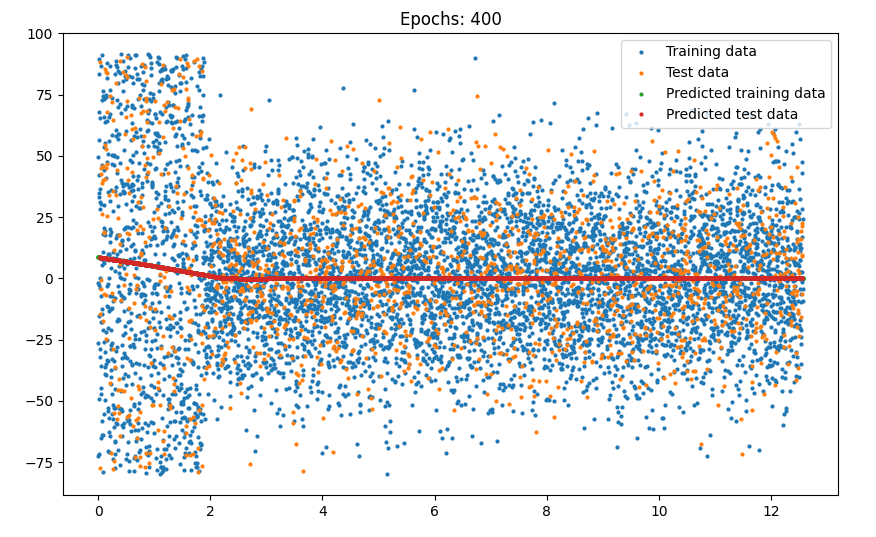
****

** Рисунок 2 Мережа тренована упродовж 200 епох**

****

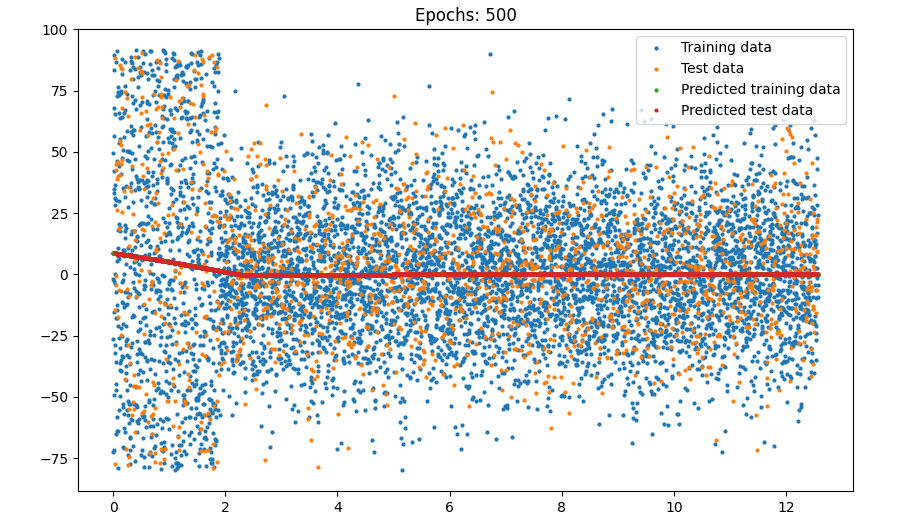
****

**Рисунок 3 Мережа тренована упродовж 300 епох**

****

****

**Рисунок 4 Мережа тренована упродовж 400 епох**

****

**Рисунок 5 Мережа тренована упродовж 500 епох**

Середньоквадратична похибка навчальної вибірки зменшується зі збільшенням кількості навчальних епох, що свідчить про те, що модель більше навчається на навчальних даних. Однак середня квадратична помилка тестового набору не демонструє послідовного покращення і може навіть почати стабілізуватися або зростати після певної кількості епох.

Зростання продуктивності на навчальній вибірці без відповідного покращення на тестовій вибірці може бути ознакою перенавчання. Перенастроювання відбувається, коли модель стає занадто спеціалізованою на навчальних даних і погано працює на нових, невідомих даних.

**ІІІ. Довести адекватність сформованих моделей та працездатність розробленого скріпта.**

**Розроблений код повинен бути раціональним та відповідати вимогам до чистого коду.**

**Частина 1**

**#**

**import numpy as np**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**from sklearn.neural\_network import MLPRegressor**

**from sklearn.model\_selection import train\_test\_split**

**from sklearn.metrics import mean\_squared\_error**

**# Step 1: Generate the dataset**

**x = np.linspace(0, 4 \* np.pi, 8000)**

**y = np.cos(x)**

**noise = np.random.normal(0, 24, y.shape)**

**y += noise**

**anomalies = np.random.uniform(np.min(y), np.max(y), int(0.15 \* len(y)))**

**y[:len(anomalies)] = anomalies**

**data = pd.DataFrame({'x': x, 'y': y})**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(data[['x']], data['y'], test\_size=0.2, random\_state=42)**

**epochs = [100, 200, 300, 400, 500]**

**for epoch in epochs:**

**model = MLPRegressor(hidden\_layer\_sizes=(100,), activation='relu', solver='adam', max\_iter=epoch)**

**model.fit(X\_train, y\_train)**

**y\_pred\_train = model.predict(X\_train)**

**y\_pred\_test = model.predict(X\_test)**

**plt.figure(figsize=(10, 6))**

**plt.plot(X\_train, y\_train, 'o', label='Training data', markersize=2)**

**plt.plot(X\_test, y\_test, 'o', label='Test data', markersize=2)**

**plt.plot(X\_train, y\_pred\_train, 'o', label='Predicted training data', markersize=2)**

**plt.plot(X\_test, y\_pred\_test, 'o', label='Predicted test data', markersize=2)**

**plt.legend()**

**plt.title(f'Epochs: {epoch}')**

**plt.show()**

**print(f'Epochs: {epoch}, Training set mean squared error: ', mean\_squared\_error(y\_train, y\_pred\_train))**

**print(f'Epochs: {epoch}, Test set mean squared error: ', mean\_squared\_error(y\_test, y\_pred\_test))**

**III. Висновок:**

виявив дослідив та узагальнив особливості застосування штучного інтелекту для аналізу даних з використанням спеціалізованих пакетів мови програмування Python.

Виконав: студент Петренко П.П.