ЗВІТ З ЛАБОРАТОРНОЇ РОБОТИ №3

за курсом "Python для Data Scientist"

студента/студентки групи ПА-

HarryJamesPotter/HermioneJeanGranger

кафедра комп’ютерних технологій, ДНУ

2024/2025

Тема: «Побудова моделі для вирішення багатофакторної задачі»

Постановка задачі:

Обрати **ОДНУ** з моделей для вирішення багатофакторної задачі:

* Модель штучного інтелекту: Нейронну мережу, Генетичний алгоритм, мурашиний алгоритм, Метод опорних векторів, Рекурентні нейронні мережі (RNN), Випадковий ліс (Random Forest), Еволюційні алгоритми або інші;
* Кластерний аналіз: методи класифікації/кластеризації, Метод К-найближчих сусідів (k-NN), інші;
* Баєсівські мережі;
* Самоорганізаційна карта Кохонена;
* Метод головних компонент (PCA);
* Методи лінійних та нелінійних регресії: Нелинейные Модели ARX та інші;
* Логістичні діаграми та рекуррентный анализ;

Обравши модель, виконати її налаштування, для вирішення **ОДНІЄЇ** з задач:

* Класифікації
* Кластеризації
* Керування
* Прогнозу
* Інші (Ідентифікації, Розпізнавання, … )

Проаналізувати точність моделі за допомогою **ОДНОГО** з методів**:**

1. **Загальна точність** (Accuracy) – відсоткове співвідношення правильно класифікованих об'єктів до загальної кількості.
2. **Матриця невідповідностей** (Confusion matrix) – таблиця, що відображає кількість правильних і помилкових класифікацій для кожного класу. <https://en.wikipedia.org/wiki/Confusion_matrix>
3. **ROC-крива (Receiver Operating Characteristic Curve)** – графік залежності True Positive Rate від False Positive Rate при різних порогах класифікації.
4. **Mean Squared Error (MSE)** – середньоквадратична похибка, більш чутлива до великих відхилень.
5. **Accuracy, Precision, Recall,** **F1-score.**
6. **PR-крива (Precision-Recall Curve)** – графік залежності Precision від Recall, особливо корисний при дисбалансі класів.
7. **Інші.**

Приклади наборів даних та ресурси:

* Seaborn Datasets: <https://github.com/mwaskom/seaborn-data>
* Kaggle Datasets: <https://www.kaggle.com/datasets>
* UCI Machine Learning Repository: <https://archive.ics.uci.edu/>
* Google Dataset Search: <https://datasetsearch.research.google.com/>
* Data.gov: <https://www.data.gov/>
* World Bank Open Data: <https://data.worldbank.org/>
* OpenML: <https://www.openml.org/>
* Amazon AWS Public Datasets: <https://aws.amazon.com/public-datasets/>
* DataIsBeautiful subreddit: <https://www.reddit.com/r/dataisbeautiful/>
* Quandl: <https://www.quandl.com/>
* GitHub: <https://github.com/awesomedata/awesome-public-datasets>
* Hugging Face: <https://huggingface.co/datasets?sort=trending>
* Титанік: <https://raw.githubusercontent.com/mwaskom/seaborn-data/master/titanic.csv>
* <https://vc.ru/u/1389654-machine-learning/581407-krupneyshie-otkrytye-datasety-dlya-mashinnogo-obucheniya>
* <https://www.linkedin.com/pulse/10-amazing-resources-access-datasets-boost-your-data-michael/>

Бібліотеки

* <https://seaborn.pydata.org/>
* PyTorch:

<https://pytorch.org/>

* Autoregressive AR-X(p) model: <https://www.statsmodels.org/dev/generated/statsmodels.tsa.ar_model.AutoReg.html>

Посилання:

* Iris Flowers Classification with PyTorch:

<https://medium.com/@ozgur.ersoz3/iris-flowers-classification-with-pytorch-cd80c8aeeb2c>

* Autoregressive (AR) Models Python Examples: Time-series Forecasting:

<https://vitalflux.com/autoregressive-ar-models-with-python-examples/>

* <https://khashtamov.com/ru/pandas-introduction/>
* <https://habr.com/ru/companies/otus/articles/741064/>
* <https://www.datacamp.com/blog/what-is-data-science-the-definitive-guide>