МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ “ЛЬВІВСЬКА ПОЛІТЕХНІКА”

Інститут КОМП’ЮТЕРНИХ НАУК ТА ІНФОРМАЦІЙНИХ технологій

кафедра систем штучного інтелекту



**Звіт**

про виконання лабораторної роботи №2

з курсу «Машинне навчання»

**Виконав**:

ст. групи КНСШ-12

Сидорук М. В.

**Перевірила**:

Бойко Н.І.

Львів – 2023

**Тема:** аналіз оптимізації процесу розглянутого методу

**Виконання завдання:**

1. **Змінити модель**

В роботі розглядалася CNN-LSTM модель. Така модель є досить великою, що впливає на час тренування а також вона буде легко перенавчатись на малих даних. Тому є сенс використати лінійну регресію або DNN.

1. **Змінити нормалізацію даних**

У випадку нормалізованих даних ми будемо передавати в модель дані в діапазоні від 0 до 1. На виході теж будуть отримані такі дані. Проте у випадку передбачення вмісту вуглецю в повітрі, фактично дані будуть лише рости. Якщо нормалізовувати дані таким чином де 0 це мінімальне у датасеті а 1 це максимальне, ми ніколи не отримаємо на вихід дані які будуть більші за максимальне у датасеті. Тому це необхідно змінити. Якщо модель лінійно-регресивна, можна забрати нормалізацію, це не сильно вплине на точність. Також можна використати більш досконалі методи нормалізації.

1. **Збільшити датасет**

Однією з проблем методу, що був реалізований раніше є те що нормалізація проводилася для одного набору даних разом. Це означає, що дані про забруденння з Китаю і дані з африканських країн були в одному діапазоні. Таким чином неможливо було зібрати хороші нормалізовані вхідні дані, оскільки для моделі не було великої різниці для цих країн, враховуючи кратну різницю (дестяки тисяч разів) між ними. Якщо забрати нормалізацію, провести окрему нормалізацію для кожної з країн або використати інший спосіб можна суттєво збільшити датасет. Це дозволить використати більш потужну модель без ризику перенавчання.

**Висновок:** в даній лабораторній роботі я навів можливі способи оптимізації алгоритму описаного у вибраній статті.