

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 №1  
з курсу «Машинне навчання»

**Виконав**

ст. групи КНСШ 12

Вовчак Л.В.

**Перевірив:**

Телішевський П.А.

ЛЬВІВ – 2023

Тема: порівняння отриманих результатів з експериментами у статті.

В обраній статті основною досліджуваною ідеєю було: Прогнозування часових рядів сезонних даних за допомогою методів машинного навчання. Реалізація відбувалась на мові програмування Python з використанням додатку Jupyter notebook.

Моделі для прогнозування часових рядів з сезонністю можуть бути використані для створення автоматичних систем у реальному часі. Наприклад, прогнозування потоку води в очисних спорудах може бути використане для розрахунку оптимального споживання електроенергії. У статті описана аналіз продуктивності різних методів машинного навчання (SARIMA, Holt-Winters Exponential Smoothing, ETS, Facebook Prophet, XGBoost та Long Short-Term Memory) і алгоритмів підготовки даних. Описано загальну методологію побудови моделі та вимоги до вихідних наборів даних. Усі моделі використовують фактичні дані від датчиків системи моніторингу. Новизна цієї роботи полягає в підході, який дозволяє використовувати обмежені історичні набори даних для отримання прогнозів з розумною точністю. Реалізовані алгоритми дозволили досягти точності R-Squared більше 0,95. Час розрахунку прогнозів мінімізований, що може бути використано для запуску алгоритму в системах реального часу та вбудованих системах.

Для замірів використовувалися наступні метрики:

***R<sup>2</sup> (коефіцієнт детермінації):*** Значення:  $R^2$  вимірює, наскільки добре модель пояснює варіацію в даних. Він приймає значення від 0 до 1, де 1 вказує на ідеальне призначення моделі, а 0 вказує на відсутність зв'язку. Інтерпретація: Високе значення  $R^2$  (наприклад, 0,8) означає, що модель добре підходить до даних, тоді як низьке значення (наприклад, 0,2) свідчить про погану відповідність.

***MSE (середньоквадратична похибка):*** Значення: MSE вимірює середньоквадратичну відстань між фактичними значеннями і прогнозами моделі. Він використовує квадрати похибок, тому великі похибки отримують більше ваги. Інтерпретація: Менше значення MSE вказує на кращу точність моделі.

***RMSE (середній квадратичний залишок)*** Значення: RMSE - це квадратний корінь середньоквадратичної похибки, яка вимірює середньоквадратичну відстань між

фактичними значеннями і прогнозами моделі. Інтерпретація: Ця метрика також вказує на точність моделі, але в тих же одиницях, що і вихідні дані.

**MAE (середньоабсолютна похибка):** Значення: MAE вимірює середню абсолютну відстань між фактичними значеннями і прогнозами моделі. Він не використовує квадрати похибок, тому всі похибки мають однакову вагу. Інтерпретація: Менше значення MAE вказує на кращу точність моделі.

**MAPE (середньоабсолютна відсоткова похибка):** Значення: MAPE вимірює середню абсолютну відсоткову похибку між фактичними значеннями і прогнозами моделі. Він виражений у відсотках. Інтерпретація: Менше значення MAPE вказує на більшу точність прогнозів моделі в відсотках.

Результати отримані в статті:

Performance Evaluation for LSTM

R2: 0.91

MSE: 0.17

RMSE: 0.42

MAE: 0.27

MAPE 0.05%

Результати отримані в моєму дослідженні:

Performance Evaluation for LTSM

R2: 0.96

MSE: 0.08

RMSE: 0.28

MAE: 0.20

MAPE: 0.03%

### **Порівняння результатів:**

Очевидним є значне покращення результатів в моєму дослідженні в порівнянні з результатами отриманими в статті. Так як моє дослідження ідентичне експериментам з статті – один із очевидних висновків полягає в тому, що оновлені версії використаних бібліотек (*pandas*, *numpy*, *keras*, *sklearn*) здатні внести суттєвий внесок у покращення результатів дослідження. Розглянемо, як саме оновлення цих бібліотек може вплинути на покращення результатів:

**Pandas:** Оновлення бібліотеки Pandas може призвести до поліпшення швидкості та продуктивності обробки даних, а також розширення функціональності для роботи з таблицями. Це сприяє більш точному та ефективному аналізу даних, що, в свою чергу, може покращити результати дослідження.

**NumPy:** Оновлення цієї бібліотеки може включати оптимізації для чисельних обчислень та роботи з масивами. Це дозволяє підвищити продуктивність обчислень та зменшити час, необхідний для обробки великих обсягів даних. У результаті це може сприяти отриманню більш точних результатів у дослідженні.

**Keras:** Оновлення Keras можуть включати нові алгоритми та покращення у тренуванні нейронних мереж. Це може позитивно вплинути на якість та швидкість навчання моделей машинного навчання, що застосовуються у дослідженні. Дуже імовірно що саме оновлена версія цієї бібліотеки чинить найбільший вплив на показники дослідження, адже саме вона містить в собі такі ключові класи:

- *Sequential* визначає модель глибокого навчання, де шари додаються один за одним в лінійному порядку. Він вказує, як будуть послідовно виконуватися шари нейронної мережі, один за іншим, від вхідних даних до виходу.
- *Dense* представляє звичайний шар нейронів. В кожен нейрон на цьому шарі підсумовуються значення від усіх нейронів на попередньому шарі, і цей підсумок передається через функцію активації.
- *LSTM* відповідний методі прогнозування часових рядів, який ми і досліджуємо, саме його оновлення мають найбільший вплив.

**Scikit-Learn (Sklearn):** Оновлення Scikit-Learn можуть вносити поліпшення у методи машинного навчання та кращий вибір моделей. Також, вони можуть додавати нові функції для валідації та оцінки моделей, що важливо для отримання більш точних результатів.

**Висновок:** Виконуючи дану лабораторну роботу, мною було порівняно результати отримані в статті та в моєму дослідженні та проаналізовано їх відмінності, оновлення певних бібліотек вплинуло на покращення продуктивності та точності дослідження, роблячи його результати більш надійними та актуальними порівняно з результатами статті.