

ЧИСЕЛЬНЕ МОДЕЛЮВАННЯ ПОГОДИ ЗА ДОПОМОГОЮ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

¹Вінницький національний технічний університет, Україна

Анотація

В роботі проведено аналіз та здійснено огляд найкращих рішень, яких було представлено на конкурсі Kaggle “Бінарне прогнозування з набором даних про опади”, що спрямований на розробку моделей машинного навчання для прогнозування ймовірності випадання опадів. Завдання полягає у бінарній класифікації: визначити, чи будуть опади (1) чи ні (0) на основі метеорологічних даних. Було виконано аналіз та опис набору даних.

Ключові слова: метеорологічні дані, інформаційні технології, машинне навчання, аналіз даних, передбачення опадів, ознаки, моделі машинного навчання.

Abstracts

This paper analyzes and reviews the best solutions submitted to the Kaggle competition “Binary Prediction with Precipitation Dataset”, which aims to develop machine learning models for predicting the probability of precipitation. The task is a binary classification: to determine whether there will be precipitation (1) or not (0) based on meteorological data. The data set was analyzed and described.

Keywords: meteorological data, information technology, machine learning, data analysis, precipitation forecasting, features, machine learning models.

Вступ

Прогнозування опадів є ключовим завданням у метеорології та має широкий спектр застосувань у різних сферах діяльності [1]. Використання даних про опади для прогнозування дозволяє запобігати природним катастрофам, наприклад, своєчасне передбачення сильних дощів та штормів допомагає протидіяти повеням, зсувам ґрунту та іншим стихійним лихам. Прогнозування опадів підказує фермерам, як планувати іригацію, посівні роботи та збирання врожаю, що зменшує втрати врожаю через посуху або надмірну вологість. Точне прогнозування випадіння дощу значно спрощує управління водними ресурсами, дозволяючи ефективніше керувати водосховищами, дамбами та гідроелектростанціями. Це забезпечує оптимальне використання води в регіонах з її нестачею або надлишком. Окрім того, детальні прогнози полегшують авіаційним і морським перевізникам уникати несприятливих погодних умов. Автомобільний транспорт може адаптувати маршрути відповідно до погодних умов. Вітрова енергетика та гідроенергетика залежать від достовірних прогнозів опадів для ефективного планування виробництва енергії, що оптимізує розрахунки навантаження на енергосистему під час екстремальних погодних умов. На додаток, передбачення дощу може запобігти поширенню хвороб, пов'язаних із застоєм води (наприклад, малярії або холери), та зменшує ризики переохолодження або теплових ударів у різних кліматичних зонах.

Результати дослідження

Набір даних для цього конкурсу (як навчальний, так і тестовий) було створено на основі моделі глибокого навчання, натренованої на наборі даних «Прогнозування опадів з використанням машинного навчання» [2]. Розподіл ознак близький до оригіналу, але не тотожний йому. Вхідні дані містять різні метеорологічні показники, такі як температура, вологість, атмосферний тиск, швидкість вітру та інші (рис. 1).

	id	day	pressure	maxtemp	temparature	mintemp	dewpoint	humidity	cloud	sunshine	winddirection	windspeed	rainfall
	0	1	1017.4	21.2	20.6	19.9	19.4	87.0	88.0	1.1	60.0	17.2	1
	1	2	1019.5	16.2	16.9	15.8	15.4	95.0	91.0	0.0	50.0	21.9	1
	2	3	1024.1	19.4	16.1	14.6	9.3	75.0	47.0	8.3	70.0	18.1	1
	3	4	1013.4	18.1	17.8	16.9	16.8	95.0	95.0	0.0	60.0	35.6	1
	4	5	1021.8	21.3	18.4	15.2	9.6	52.0	45.0	3.6	40.0	24.8	0
	...	...	...	...	...	...	...	...	...	...	...	...	...
85	2185	361	1014.6	23.2	20.6	19.1	19.9	97.0	88.0	0.1	40.0	22.1	1
86	2186	362	1012.4	17.2	17.3	16.3	15.3	91.0	88.0	0.0	50.0	35.3	1
87	2187	363	1013.3	19.0	16.3	14.3	12.6	79.0	79.0	5.0	40.0	32.9	1
88	2188	364	1022.3	16.4	15.2	13.8	14.7	92.0	93.0	0.1	40.0	18.0	1
89	2189	365	1013.8	21.2	19.1	18.0	18.0	89.0	88.0	1.0	70.0	48.0	1

2190 rows × 13 columns

Рисунок 1. Приклад метеорологічних показників, що містить набір даних

Нижче наведено кілька ключових спостережень та висновків із кореляційної матриці числових ознак, яка демонструє ступінь лінійної залежності між цими змінними (рис. 2).

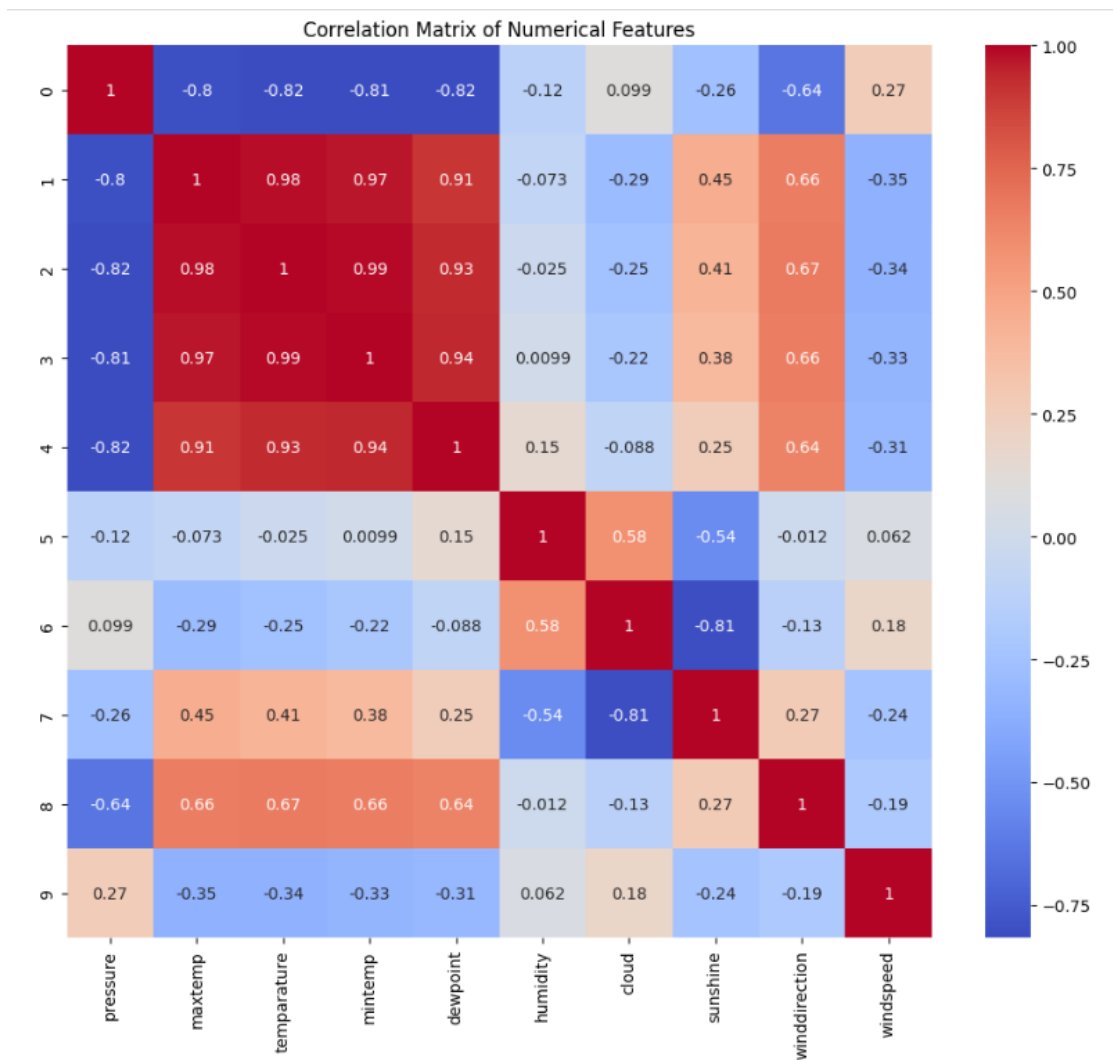


Рисунок 2. Кореляційна матриця числових ознак

Матриця кореляцій підтверджує типові метеорологічні закономірності: висока кореляція між температурними показниками, протилежний зв'язок між температурою та тиском, а також сильна залежність між хмарністю та сонячним сйвом.

Задача передбачення опадів належить до методів навчання з учителем, оскільки ми маємо наявні дані з відомими мітками (чи були опади, чи ні). Це задача “класифікації”, тому що цільова змінна (наявність опадів) набуває “дискретних значень”. У конкурсі використовується метрика “AUC-ROC (Area Under the Receiver Operating Characteristic Curve)”, яка вимірює здатність моделі розділяти класи, враховуючи співвідношення між справжніми позитивами та хибними позитивами при різних порогах класифікації. Для визначення опадів необхідно створити, налаштувати та провести тренування моделей машинного навчання. На конкурсі було застосовано серед найбільш релевантних різні алгоритми класифікації, такі як логістична регресія, дерева рішень, ансамблеві методи (Random Forest, Gradient Boosting) або нейронні мережі [3]. Результат тренування моделей відображено в точності передбачення цільової ознаки (опадів) за AUC-ROC(рис. 3).

Model	AUC-ROC
MLP (Multi-Layer Perceptron)	0.88120
Нейронна мережа FeedForward з тюнером Keras	0.87476
LogisticRegression	0.87396
NN на базі LSTM	0.86296
AutoML (LightGBM, CatBoost, XGBoost, Random Forest та інші)	0.83963
XGBoost	0.83614
CP-SAT Ensemble	0.86108
Decision Trees	0.83427
LightGBM with Bayesian Optimization	0.86403

На рисунку 3 видно, що найкращою моделлю є багатошаровий персептрон (MLP) за критерієм якості 0.88120 (чим ближче значення до 1, тим краща точність). Взагалі в усіх моделей було досягнуто достатньої точності передбачення не лише завдяки вдалому вибору та удосконаленні обраної моделі, а й завдяки коректному аналізу та обробці вхідних даних та інжинірингу ознак.

Висновки

Прогнозування опадів є критично важливим для економіки, безпеки та добробуту суспільства. Використання сучасних алгоритмів машинного навчання для аналізу метеорологічних даних дозволяє значно підвищити точність прогнозів і мінімізувати негативні наслідки погодних явищ.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Кліматичні ризики функціонування галузей економіки України в умовах зміни клімату: монографія / за ред. С. М. Степаненка, А. М. Польового. Одеський державний екологічний університет. Одеса: ТЕС, 2018. 548 с.
2. Binary Prediction with a Rainfall Dataset, Kaggle, 2025 [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/competitions/playground-series-s5e3/overview>
3. Наука про дані: машинне навчання та інтелектуальний аналіз даних : [Електронний ресурс] / В. Б. Мокін, М. В. Дратований – Вінниця: ВНТУ, 2024. – 258 с. URL: https://iq.vntu.edu.ua/method/getfile.php?fname=150036.pdf&x=1&card_id=64543&id=150036

Лісачов Роман Всеволодович – студент групи 2ІСТ-24м, Факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: lisachovroman@gmail.com;

Науковий керівник: **Мокін Віталій Борисович** – д-р. техн. наук, проф., завідувач кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: ybmokin@vntu.edu.ua

Lisachov Roman V. - student of group 2IST-24m, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: lisachovroman@gmail.com

Scientific adviser: **Mokin Vitalii B.** – Dr. tech. Sciences, Prof., Head of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: ybmokin@vntu.edu.ua