



О. М. Шопський, О. В. Хлевной

Львівський державний університет безпеки життєдіяльності, м. Львів, Україна

ORCID: <https://orcid.org/0009-0000-1761-6643> – О. М. Шопський

<https://orcid.org/0000-0003-2846-3480> – О. В. Хлевной



orest.shopsky@gmail.com

МЕТОД ПРОГНОЗУВАННЯ КІЛЬКОСТІ ПОЖЕЖ В ЖИТЛОВОМУ СЕКТОРІ НА ОСНОВІ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ

Пожежі в житловому секторі є серйозною загрозою для безпеки населення. Вони спричиняють значні збитки та потребують ефективних методів прогнозування для їх профілактики. У цій статті представлено метод прогнозування кількості пожеж у житловому секторі на основі алгоритмів машинного навчання, який базується на аналізі даних про виїзди рятувальних підрозділів Львівського району. Дані включають інформацію про час, місце й характер подій, агреговану за шестикутниками площею 3 км². Як предиктори використано густоту населення (осіб/км²) і місяць року, закодований як набір із 11 dummy-змінних для врахування сезонності. Для тренування моделей обрано дані за 5 років (2016-2020), для тестування використовувалися дані за 2021 рік. Загальний обсяг даних перевищив 3 тисячі записів, що забезпечує надійну основу для аналізу.

У дослідженні використано чотири моделі: лінійну регресію, поліноміальну регресію (другого степеня), випадковий ліс і нейронну мережу (багатошаровий перцептрон). Лінійна регресія дала змогу оцінити базовий лінійний зв'язок між предикторами та кількістю пожеж, тоді як поліноміальна регресія дозволила врахувати нелінійні ефекти, такі як квадратичний вплив густоти. Випадковий ліс створив передумови для моделювання складних взаємодій між змінними, а нейронна мережа дала можливість виявити приховані закономірності. Точність оцінено за метриками MAE, RMSE і R^2 на тренувальній вибірці. Найкращі результати показав випадковий ліс із найнижчими похибками, що свідчить про його перевагу в задачах із вираженою сезонністю й середнім обсягом даних. Нейронна мережа виявилася чутливою до обсягу вибірки, а недоліками лінійної та поліноміальної регресій стала обмежена гнучкість. Модель випадкового лісу також було застосовано для прогнозування кількості пожеж на тестових даних. Порівняння прогнозів із реальними значеннями показало задовільну точність, хоча відхилення в зимові місяці вказують на потребу врахування додаткових факторів (наприклад, погодних умов). Метод дозволяє ідентифікувати зони підвищеного ризику, що може бути використано для перерозподілу ресурсів підрозділів цивільного захисту і планування профілактичних заходів. Перспективи включають додавання температури, опадів і просторових моделей, таких як GWR, для підвищення точності.

Ключові слова: прогнозування пожеж, житловий сектор, машинне навчання, лінійна регресія, поліноміальна регресія, випадковий ліс, нейронна мережа.

О. М. Shopskyi, O. V. Khlevnoi

Lviv State University of Life Safety, Lviv, Ukraine

THE METHOD OF FORECASTING THE NUMBER OF RESIDENTIAL SECTOR FIRES BASED ON MACHINE LEARNING ALGORITHMS

Fires in residential buildings pose a serious threat to public safety, causing significant damage and necessitating effective forecasting methods for their prevention. This article presents a method for predicting the number of residential sector fires using machine learning algorithms, based on an analysis of data from emergency response unit deployments in the L'viv District. The data encompass information on the timing, location, and nature of incidents, aggregated into hexagons with an area of 3 km². Population density (persons/km²) and the month of the year, encoded as a set of 11 dummy variables to account for seasonality, were used as predictors. Data for training the models were drawn from a 5-year period (2016–2020), while data from 2021 were used for testing. The total dataset exceeded 3,000 records, providing a robust foundation for analysis.

The study employed four models: linear regression, polynomial regression (second degree), random forest, and a neural network (multilayer perceptron). Linear regression enabled an assessment of the baseline linear relationship

between predictors and the number of fires, while polynomial regression accounted for nonlinear effects, such as the quadratic impact of population density. Random forest facilitated the modeling of complex interactions between variables, and the neural network offered the potential to uncover hidden patterns. Accuracy was evaluated using MAE, RMSE, and R^2 metrics on the training dataset. The random forest model yielded the best results, with the lowest error rates, highlighting its superiority in tasks involving pronounced seasonality and moderate data volumes. The neural network was sensitive to sample size, while linear and polynomial regressions were limited by their lack of flexibility. The random forest model was also applied to predict fire occurrences in the test data. A comparison of predictions with actual values demonstrated satisfactory accuracy, though deviations during winter months suggest the need to incorporate additional factors, such as weather conditions. This method enables the identification of high-risk zones, which can inform the reallocation of civil protection unit resources and the planning of preventive measures. Future improvements could involve integrating temperature, precipitation, and spatially explicit models like GWR to enhance prediction accuracy.

Key words: fires forecasting, residential buildings, machine learning, linear regression, polynomial regression, random forest, neural network

Постановка проблеми. Пожежі в житловому секторі становлять одну з ключових загроз для безпеки населення, спричиняючи значні людські жертви, економічні збитки та створюючи додаткове навантаження на рятувальні служби. За даними Державної служби України з надзвичайних ситуацій, щороку в Україні фіксуються десятки тисяч пожеж, значна частка яких припадає саме на житлові об'єкти. Ці події мають виражену сезонну залежність, пов'язану з погодними умовами, опалювальним періодом чи людською діяльністю, а також просторову неоднорідність, зумовлену такими факторами, як густота населення та тип забудови. Незважаючи на великий обсяг накопичених даних про надзвичайні ситуації, їх аналіз і використання для прогнозування залишаються недостатньо розвиненими через їхню неструктурованість, яка ускладнює ефективну профілактику та реагування.

Сучасні підходи до прогнозування пожеж часто базуються на статистичних методах або експертних оцінках, які повною мірою не враховують складні взаємозв'язки між факторами ризику. Наприклад, традиційні моделі можуть ігнорувати нелінійний вплив густоти населення, коли зростання кількості жителів у певній зоні непропорційно підвищує ймовірність виникнення пожеж через щільність інфраструктури чи порушення правил безпеки. Аналогічно, сезонні коливання, потребують детальнішого аналізу, адже такі чинники, як суха погода влітку чи використання опалювальних приладів узимку, мають різний ефект у різних регіонах. Отже, виникає потреба в розробці методу, який би враховував ці особливості й забезпечував точне прогнозування кількості пожеж на локальному рівні.

У цьому контексті значний потенціал мають дані про виїзди рятувальних підрозділів Львівської області, що накопичувалися більше десяти років. Ці записи містять детальну інформацію про час, місце й характер подій, зокрема пожеж у житловому секторі, і можуть слугувати основою

для побудови прогностичних моделей. Проте дані мають неструктурований характер і потребують систематизації, агрегації та обробки для використання в аналітиці. Наприклад, агрегація за просторовими одиницями – шестикутниками площею 3 км² – дозволяє врахувати географічну специфіку, тоді як включення таких предикторів, як густота населення та місяць року, дає змогу моделювати просторово-часові закономірності.

Проблема полягає в тому, що традиційні методи аналізу, такі як лінійна регресія, можуть бути недостатньо гнучкими для виявлення складних залежностей у таких даних, тоді як сучасні алгоритми машинного навчання, наприклад ансамблеві методи чи нейронні мережі, ще недостатньо застосовувалися до цієї задачі в локальному контексті. Крім того, відсутність порівняльного аналізу різних підходів ускладнює вибір оптимального методу для практичного використання. Таким чином, існує потреба в розробці й оцінці методу прогнозування кількості пожеж у житловому секторі, який би базувався на алгоритмах машинного навчання, використовував історичні дані про виїзди та враховував вплив густоти населення й сезонності.

Аналіз останніх досліджень і публікацій. Прогнозування пожеж у житловому секторі є важливою задачею, що привертає увагу дослідників у галузях комп'ютерних наук, безпеки та аналізу даних. Значний обсяг літератури присвячено аналізу факторів ризику та розробці моделей для передбачення надзвичайних ситуацій, однак використання алгоритмів машинного навчання для прогнозування кількості пожеж на основі історичних даних про виїзди рятувальних підрозділів залишається відносно новим напрямом.

Одним із ключових підходів до прогнозування пожеж є статистичний аналіз. Наприклад, у роботі [1] використано регресійні моделі для оцінки впливу погодних умов на поширення лісових пожеж в Житомирській області. У роботі [2] досліджено вплив погодних умов на частоту виникнення пожеж в природних екосистемах. Автори показали, що лінійна регресія може

ефективно виявляти базові закономірності, однак її точність знижується за наявності нелінійних залежностей. Аналогічно, у [3] використано генетичні алгоритми та нейронні мережі для дослідження кількості пожеж у Великобританії, враховуючи сезонність і тип забудови. Ці дослідження підкреслюють важливість сезонності як предиктора, що узгоджується з запропонованим у цій роботі підходом, хоча автори не використовували просторову агрегацію за рівновеликими фрагментами території.

Розвиток геоінформаційних систем (ГІС) сприяв появі просторово орієнтованих методів. Так, у роботі [4] запропоновано географічно зважену регресію (GWR) для прогнозування ризиків лісових пожеж. Автори доводять, що локальні моделі переважають глобальні за точністю в умовах просторової неоднорідності. Хоча їхня робота зосереджена на пожежах в природних екосистемах, а не в житловому секторі, концепція просторової деталізації є релевантною і для наших даних, агрегованих за шестикутниками. Водночас GWR рідко застосовується до міських середовищ із високою густотою населення, що вказує на прогалину в літературі.

Останніми роками алгоритми машинного навчання набули популярності в прогнозуванні надзвичайних ситуацій. У дослідженнях [5-8] випадковий ліс використано для прогнозування кількості лісових пожеж на основі кліматичних та інших факторів. У всіх випадках моделі продемонстрували високу точність, що підкреслює переваги ансамблевих методів у задачах із нелінійними залежностями. Подібним чином у роботах [9-11] було застосовано нейронні мережі для аналізу часових рядів пожеж, враховуючи сезонність і кліматичні фактори. Їхні результати свідчать про здатність нейронних мереж виявляти складні патерни, однак потребу у великих обсягах даних для навчання часто згадують як обмеження.

Деякі роботи фокусуються на даних про виїзди рятувальних служб. Наприклад, у [12] проаналізовано дані про пожежі в Україні, використовуючи кластерний аналіз для ідентифікації зон ризику. Хоча наведений підхід не передбачав прогнозування кількості подій, він демонструє цінність таких даних для локального аналізу. Загалом в контексті України дослідження пожеж із використанням машинного навчання обмежені, що залишає простір для подальших досліджень.

Аналіз літератури дозволив виявити кілька прогалин. По-перше, більшість робіт зосереджені на ймовірності виникнення пожеж (бінарний результат), а не на їхній кількості, що обмежує можливості прогнозування інтенсивності подій.

По-друге, хоча густота населення та сезонність часто розглядаються як фактори ризику, їх комбінований вплив у міських умовах із використанням даних про виїзди недостатньо вивчений. По-третє, порівняльний аналіз різних моделей машинного навчання (лінійних, нелінійних, ансамблевих і нейронних) рідко застосовується до цієї задачі, що ускладнює вибір оптимального підходу. Нарешті, локальний контекст, як-от специфіка житлового сектору Львівської області, залишається поза увагою як міжнародних, так і вітчизняних досліджень.

Отже, незважаючи на значний прогрес у прогнозуванні пожеж, існує потреба в розробці методу, який би поєднував аналіз історичних даних про виїзди, просторову агрегацію та алгоритми машинного навчання для передбачення кількості пожеж у житловому секторі. Ця стаття спрямована на вирішення цих проблем шляхом встановлення залежності кількості пожеж в житловому секторі від густоти населення та сезонності за допомогою різних алгоритмів: лінійної регресії, поліноміальної регресії, випадкового лісу та нейронної мережі. Такий підхід не лише дає змогу розширити теоретичну базу, але й пропонує практичні рішення для подальшої профілактики пожеж у житловому секторі.

Мета і завдання досліджень. Метою цього дослідження є розробка методу прогнозування кількості пожеж у житловому секторі з використанням алгоритмів машинного навчання на основі даних про виїзди рятувальних підрозділів за 2016-2021 роки. Метод має враховувати вплив густоти населення та сезонності (вираженої через місяці року) як ключових предикторів.

Для досягнення поставленої мети визначено такі **завдання дослідження**:

- Підготувати дані про виїзди рятувальних підрозділів Львівської області за 2016-2021 роки, виділити пожежі в житловому секторі, агрегувати їх за фрагментами однакової площі і сформувати набір даних із предикторами – густотою населення та місяцем (dummy-змінні), розподіливши на тренувальну (2016-2020 роки) і тестову вибірки (2021 рік).

- Розробити різні моделі машинного навчання для прогнозування кількості пожеж, зокрема, лінійну регресію, поліноміальну регресію, випадковий ліс і нейронну мережу та порівняти їх точність.

- Здійснити апробацію моделі з найвищою точністю шляхом прогнозування кількості пожеж для різних зон Львівського району та візуалізації отриманих результатів.

Методи досліджень. Для розробки методу прогнозування кількості пожеж у житловому

секторі на основі алгоритмів машинного навчання необхідно було отримати релевантні дані, які б відображали частоту та характеристики надзвичайних подій. З цією метою використано дані про всі виклики, зафіксовані в Системі оперативно-диспетчерського управління [13] Львівської області за період із 2016 по 2021 роки. Цей масив включав 131095 записів, що містили інформацію про час, місце (адреси та координати), зміст повідомлення, опис події та її класифікацію. Дані були неструктурованими через неоднорідність записів: значна кількість містила орфографічні помилки, некоректну класифікацію, а також включала виїзди, не пов'язані з реальними пожежами (навчання, господарські роботи). Така специфіка створювала перешкоди для їхнього прямого використання в моделюванні, тому знадобилася багатоступенева обробка.

Першим етапом стала систематизація даних у базі PostgreSQL. Для повнотекстового пошуку та обробки тексту встановлено Великий електронний словник української мови (ВЕСУМ) [14], який містить 426814 словоформ і підтримує нормалізацію тексту до називного відмінка [15]. Аналіз показав, що 18% записів містили помилки типу "затимління" замість "затимлення", "пжежа" замість "пожежа" чи запозичення з інших мов. Для їх усунення розроблено модуль у Python із використанням бібліотеки Hunspell через обгортку CyHunspell. Hunspell перевіряв орфографію за словником ВЕСУМ і спеціальним додатком із 1 528 термінів пожежно-рятувальної тематики [16]. Помилки виправлялися автоматично, обираючи перший варіант із запропонованих. У 85% випадків виправлення було успішним, що підвищило якість тексту для подальшого аналізу. Стоп-слова (наприклад, "та", "в") видалялися за списком із 342 одиниць, зменшивши обсяг тексту на 15%.

За допомогою SQL-запитів із модулем `to_tsvector` відфільтровано усі виїзди на пожежі, виключивши 21 категорію хибних і навчальних подій (наприклад, "Пожежно-тактичні навчання", "Хибний виклик", «Перевірка боєздатності» тощо), а також надзвичайні події, не пов'язані з пожежами (дорожньо-транспортні пригоди, замінування, промислові аварії та нещасні випадки, виявлення вибухонебезпечних предметів тощо). Це дало змогу отримати 15706 записів про виїзди на пожежі, серед яких лише 5925 були пов'язані із житловим сектором.

Ще однією проблемою стала географічна прив'язка даних до шестикутників площею 3 км². Із загальної кількості записів лише 68% мали точні координати подій, тоді як решта містила лише текстові адреси. Для їх конвертації було використано бібліотеку `geopy` у Python, яка геокодувала адреси через OpenStreetMap із

точністю 92%, що дозволило прив'язати усі записи до шестикутників в межах Львівського району (варто зауважити, що Львівський район входить до складу Львівської області і ці поняття не можна ототожнювати). У підсумку сформовано просторовий набір даних із 3211 геокодованих подій, агрегованих за 1310 шестикутниками Львівського району і більшість цих шестикутників припадала на місцевість із нульовою або низькою щільністю населення.

Кількість пожеж у житловому секторі залежно від шестикутника змінювалася у діапазоні від 0 до 19 на місяць. Додатково було введено предиктори: густота населення (осіб/км², від 1 до 15045) і *dummy*-змінні для 11 місяців (січень – базовий).

Для підготовки до моделювання дані розподілено на тренувальну вибірку (2016–2020, 60 місяців, 2657 записів) і тестову (2021, 12 місяців, 554 записи). Остаточний набір містив 3211 пожеж, агреговані за 341 шестикутником, із трьома стовпцями: кількість пожеж, густота населення, місяць.

Результати дослідження

Аналіз проводили за допомогою 4 моделей машинного навчання – лінійною регресією, поліноміальною регресією, випадковим лісом і нейронною мережею – для порівняння їх ефективності.

Дані були попередньо нормалізовані для забезпечення коректності аналізу. Густота населення стандартизована за допомогою зперетворення $z = \frac{x - \mu}{\sigma}$ (де x – поточне значення щільності, $\mu = 963$ – середнє значення, $\sigma = 2286$ – середньоквадратичне відхилення), що перевело її в діапазон із середнім 0 і стандартним відхиленням 1. Кількість пожеж залишилася ненормалізованою як цілочислова змінна. Місяць закодовано як 11 бінарних змінних (наприклад, "Липень" = [0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0], січень – базовий). Усі обчислення виконано в Google Colab із бібліотеками `scikit-learn` і `TensorFlow`.

Для попереднього аналізу зв'язків між змінними – кількістю пожеж у житловому секторі, густотою населення та місяцем року – було побудовано теплову карту кореляцій. Це дозволило виконати оцінку ступеня лінійної залежності між предикторами та цільовою змінною, а також виявити можливу мультиколінеарність між предикторами, що може вплинути на інтерпретацію моделей машинного навчання, зокрема лінійної регресії. Теплова карта дозволила візуально представити матрицю кореляцій, де значення коефіцієнта кореляції Пірсона набували значень від -1 (сильна негативна залежність) до 1 (сильна позитивна залежність), а 0 вказував на відсутність лінійного зв'язку.

Матрицю кореляцій (рисунок 1) було отримано за допомогою бібліотеки pandas у Python (функція `corr()`), яка використовує формулу Пірсона. Для візуалізації використано

бібліотеку seaborn (функція `heatmap`), із параметрами: теплова шкала від -1 до 1 (колірна схема "coolwarm"), відображення значень коефіцієнтів у клітинках.

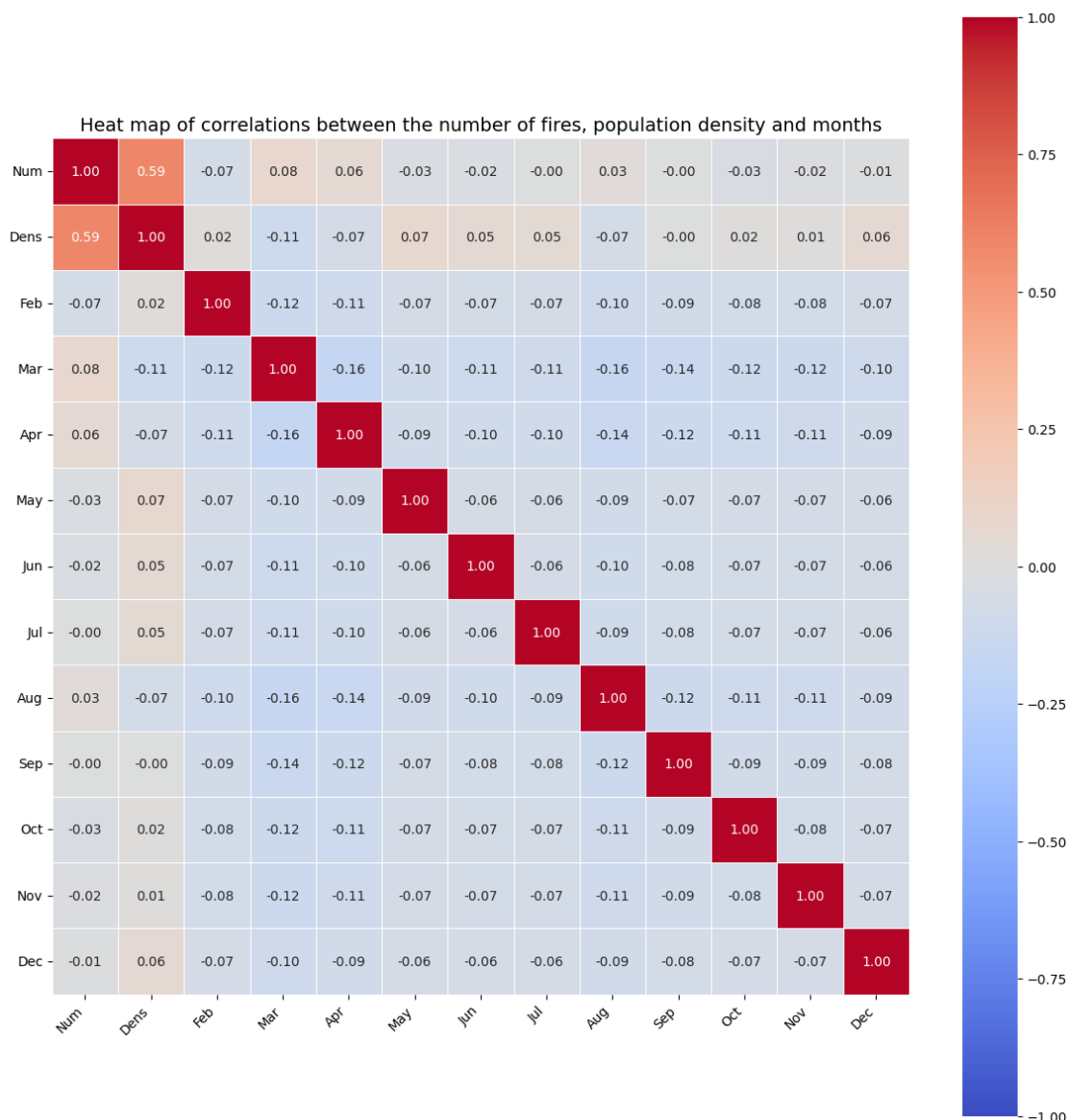


Рисунок 1 – Теплова карта кореляцій

Аналіз матриці кореляцій показав помірно сильну позитивну залежність кількості пожеж у житловому секторі від густоти населення ($r=0,591$), що підкреслює її ключову роль як предиктора. Натомість зв'язок із місяцями року виявився слабким ($|r|<0,1$), із максимальними значеннями для березня ($r=0.083$) та квітня ($r=0.058$), що може свідчити про нелінійний характер сезонності, не відображений лінійною кореляцією. Відсутність мультиколінеарності між предикторами ($|r|<0.17$) підтвердила їхню придатність для моделювання. Отримані результати вказують на доцільність застосування нелінійних моделей, таких як випадковий ліс або

нейронна мережа, для глибшого аналізу сезонних і просторових закономірностей.

Лінійна регресія. Першою для аналізу залежності кількості пожеж у житловому секторі (**Num**) від густоти населення (**Dens**) і місяців року (**Feb–Dec**) було застосовано модель лінійної регресії. Цей метод обрано як базовий для оцінки лінійних зв'язків між змінними, виявлених за результатами кореляційного аналізу. Модель реалізована в Python із бібліотекою scikit-learn (LinearRegression). Навчання проводилося на навчальній вибірці.

Коефіцієнти моделі лінійної регресії наведені у таблиці 1.

Таблиця 1
Коефіцієнти лінійної регресії

Змінна	Коефіцієнт
Dens	1,313609
Feb	0,365202
Mar	2,480012
Apr	2,177202
May	0,344747
Jun	0,615145
Jul	0,861080
Aug	1,952063
Sep	1,352198
Oct	0,871007
Nov	1,040287
Dec	0,662687
Вільний член	0,6663

Лінійна регресія продемонструвала низьку продуктивність у прогнозуванні кількості пожеж, із середньою абсолютною похибкою (MAE) 1,486, що вказує на відхилення прогнозів у середньому на 1,5 одиниці від реальних значень, середньоквадратичною похибкою (RMSE) 2,412, що підкреслює наявність більших помилок у деяких випадках, і коефіцієнтом детермінації (R^2) 0,412, який свідчить, що модель пояснює лише 41,2% варіації даних, відображаючи її обмежену здатність

вловлювати складні залежності між густотою населення, місяцями та кількістю пожеж.

Поліноміальна регресія другого степеня. Наступною для врахування можливої нелінійної залежності кількості пожеж у житловому секторі (Num) від густоти населення (Dens) і місяців року (Feb–Dec) було застосовано поліноміальну регресію другого степеня. Цей підхід розширив лінійну модель, додавши квадратичні члени та взаємодії між предикторами, що дало змогу краще моделювати складні закономірності, які не вловлюються лінійною регресією. Поліноміальні ознаки другого степеня (включаючи $Dens^2$, Feb^2 , $Dens \cdot Feb$ тощо) згенеровано за допомогою PolynomialFeatures із бібліотеки scikit-learn, що дало 90 ознак (12 початкових + їх квадрати та взаємодії).

Поліноміальна регресія показала дещо кращу продуктивність порівняно з лінійною моделлю, із середньою абсолютною похибкою MAE = 1,412, середньоквадратичною похибкою RMSE = 2,342 і коефіцієнтом детермінації $R^2 = 0,445$, який засвідчив дуже обмежене покращення завдяки нелінійним членам, але недостатню здатність повною мірою моделювати складні залежності між густотою населення, місяцями та кількістю пожеж.

Випадковий ліс. Модель RandomForest Regressor було налаштовано з такими параметрами: 100 дерев, максимальна глибина 10, random_state=42. Навчання проводилося на навчальній вибірці. Внесок кожної ознаки (Dens, Feb, ..., Dec) у прогнозування кількості пожеж за цією моделлю відображено за допомогою Bar Plot важливості ознак (рисунок 2).

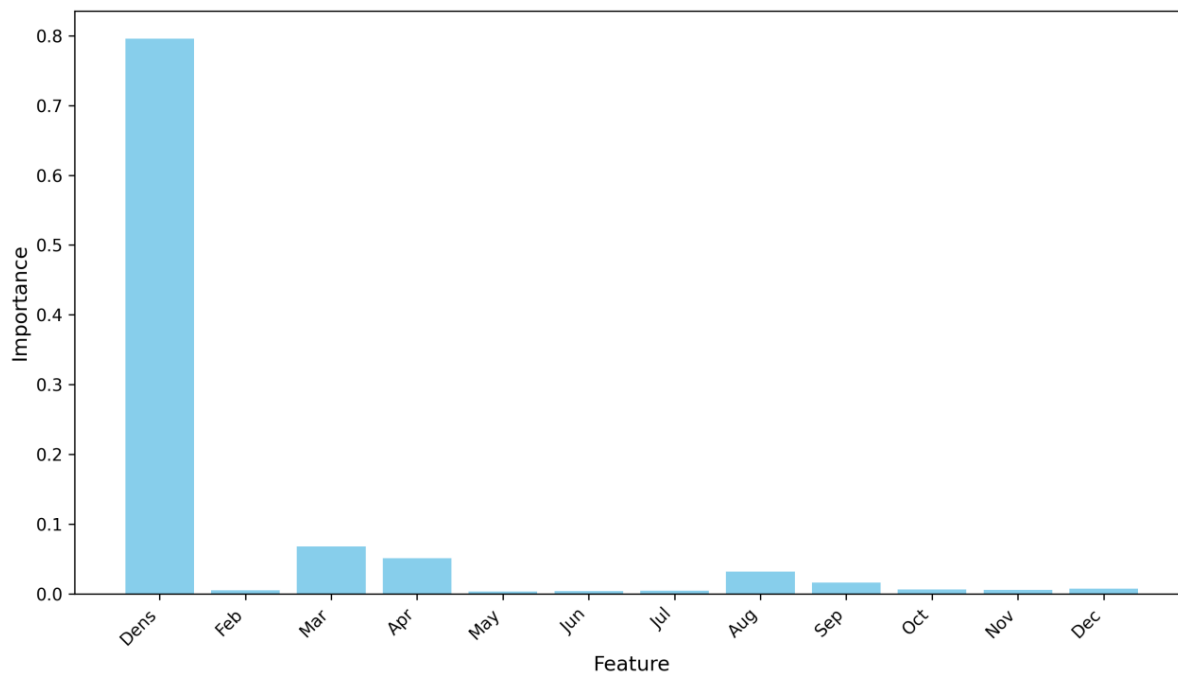


Рисунок 2 – Важливість ознак у моделі випадкового лісу

Випадковий ліс виявився дуже ефективним у прогнозуванні кількості пожеж: коефіцієнт детермінації ($R^2 = 0,841$) свідчить, що модель пояснює 84.1% варіації даних, середня абсолютна похибка $MAE = 0,855$ означає, що прогнози відхиляються від реальних значень у середньому менш ніж на 1 одиницю (для середньомісячної кількості пожеж це задовільна точність), а середньоквадратична похибка $RMSE = 1,253$ підтверджує стабільність моделі, демонструючи її здатність добре враховувати нелінійні зв'язки між густотою населення, місяцями та кількістю пожеж.

Нейронна мережа. Для регресійного аналізу за допомогою штучної нейронної мережі було застосовано багатошаровий перцептрон (MLP), що є однією з базових архітектур нейронних мереж. Така мережа здатна вловлювати нелінійні

та складні взаємодії між предикторами, що може бути перевагою порівняно з лінійними моделями чи навіть випадковим лісом.

Нейронну мережу було реалізовано через `MLPRegressor` із бібліотеки `scikit-learn` із двома прихованими шарами (32 і 16 нейронів), функцією активації `ReLU`, оптимізатором `Adam`, швидкістю навчання 0.001 та максимальною кількістю 100 ітерацій. Навчання проводилося на навчальній вибірці та продемонструвало посередні результати, а саме коефіцієнт детермінації $R^2 = 0,450$, середню абсолютну похибку $MAE = 1,393$ та середньоквадратичну похибку $RMSE = 2,334$. Ці метрики демонструють обмежену здатність ефективно моделювати складні залежності між густотою населення, місяцями та кількістю пожеж у порівнянні з іншими підходами.

Обговорення результатів досліджень. Порівняння результатів регресійного аналізу наведено у таблиці 2.

Таблиця 2

Основні метрики моделей регресії

Модель	MAE	RMSE	R^2
Лінійна регресія	1,486	2,412	0,412
Поліноміальна регресія	1,412	2,342	0,445
Випадковий ліс	0,855	1,253	0,841
Нейронна мережа	1,393	2,334	0,45

Аналіз цих результатів свідчить, що лінійна регресія показала найгірші результати серед усіх моделей. Її середня абсолютна похибка в 1,486 одиниць означає, що середнє відхилення прогнозу від реальних значень становить майже півтори одиниці, що є суттєвим для даних із діапазоном 0–19. Значення $RMSE$ (2,412) є значно вищим, ніж MAE , а отже модель допускає великі помилки в окремих випадках, ймовірно, через нездатність впоратися з вищими значеннями чи аномаліями. Коефіцієнт детермінації 0,412 вказує на слабку лінійну кореляцію місяців із кількістю пожеж (<0.1) і часткову кореляцію з густотою (0,591). Найімовірніше, причиною такого слабого результату є те, що лінійна модель припускає просту залежність, яка не відображає складні нелінійні взаємодії між предикторами, особливо коли сезонність виявляється неочевидною в лінійному контексті.

Поліноміальна регресія другого степеня має дещо вищу точність порівняно з лінійною моделлю. Її MAE 1.412 показує трохи меншу середню похибку, $RMSE$ 2,342 також нижчий, що вказує на зменшення великих відхилень, хоча вони все ще присутні. Коефіцієнт детермінації 0,445 демонструє лише незначне покращення, яке відображає внесок квадратичних членів і взаємодій, однак ці показники все ще далекі від

ідеальних, і причина, ймовірно, полягає в тому, що степінь 2 недостатній для повного моделювання нелінійності в даних. Крім того, слабкий вплив місяців, який ми бачили в кореляційному аналізі, обмежує здатність поліноміальної моделі суттєво перевершити лінійну, адже додаткові ознаки не додають значної пояснювальної сили.

Випадковий ліс, навпаки, продемонстрував дуже хорошу продуктивність. Значення $MAE = 0,855$ – це вдвічі менша похибка порівняно з іншими моделями, що робить прогнози значно точнішими, особливо для низьких значень кількості пожеж, які, судячи з середнього значення, домінують у вибірці (рисунок 3). $RMSE = 1,253$ також набагато нижчий, а менша різниця між MAE і $RMSE$ (0,398) свідчить про стабільність моделі навіть у присутності викидів чи складних випадків. Достатньо високий коефіцієнт кореляції $R^2 = 0,841$ є найкращим результатом з усіх отриманих. Така висока продуктивність пояснюється ансамблевим підходом: модель будує багато дерев, кожне з яких враховує різні аспекти даних, і їхнє усереднення дозволяє вловити нелінійні залежності та взаємодії між густотою населення і місяцями, які ігнорують простіші моделі. Домінування густоти населення (важливість $\approx 0,65$ за попередніми оцінками) і слабший, але все ж наявний внесок

місяців через нелінійні ефекти ідеально підходять для цього методу.

Нейронна мережа показала посередній результат, трохи кращий за лінійну регресію, але гірший за випадковий ліс. Значення $MAE = 1,393$ і $RMSE 2,334$ схожі на показники поліноміальної моделі та вказують на схильність до значних відхилень в окремих випадках. Коефіцієнт

детермінації $0,450$ є низьким. Зміна кількості нейронів у шарах та збільшення кількості ітерацій практично не вплинули на метрики точності. Причина такого результату, ймовірно, криється в недостатньому для ефективного навчання розмірі тренувальної вибірки, оскільки нейромережі зазвичай потребують більше даних, ніж ансамблеві методи типу випадкового лісу.

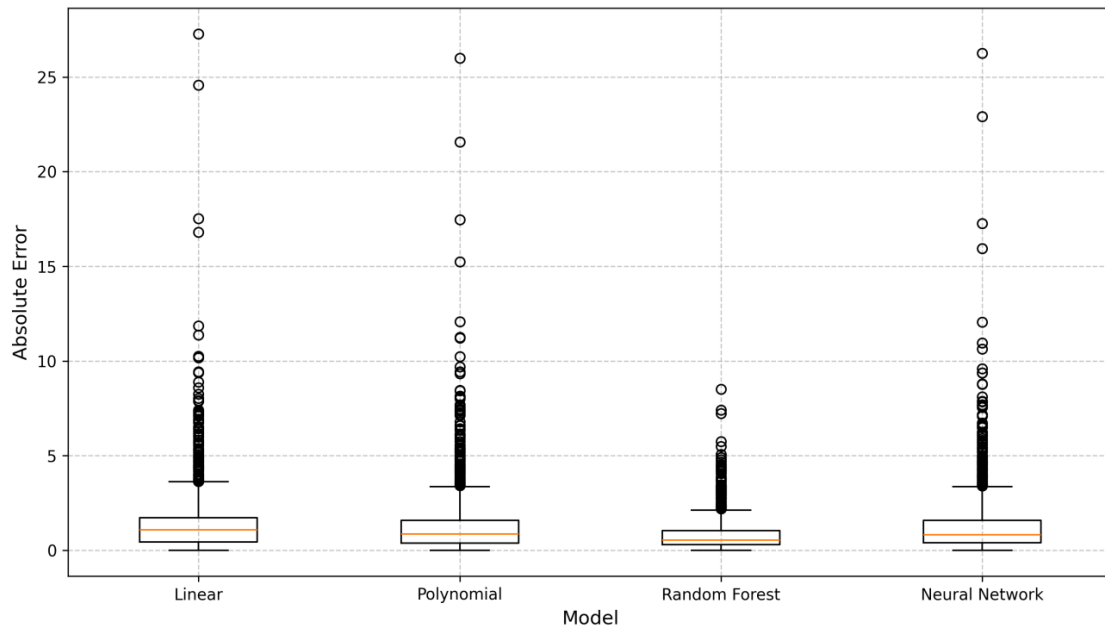


Рисунок 3 – Розподіл похибок прогнозування

Таким чином, можна зробити висновок, що прогнозувальна здатність (рисунок 4) лінійної регресії виявилась найслабшою через простоту моделі, яка не відповідає складності даних. Поліноміальна регресія додала трохи нелінійності, але її покращення було мінімальним через обмежений степінь і слабкий внесок місяців.

Нейронна мережа могла б показати кращий результат із більшою вибіркою, але в поточному вигляді вона поступається за всіма метриками. Випадковий ліс, завдяки своїй гнучкості й ансамблеві природі, значно перевершив усі інші підходи, забезпечуючи найнижчу похибку і найвищу точність.

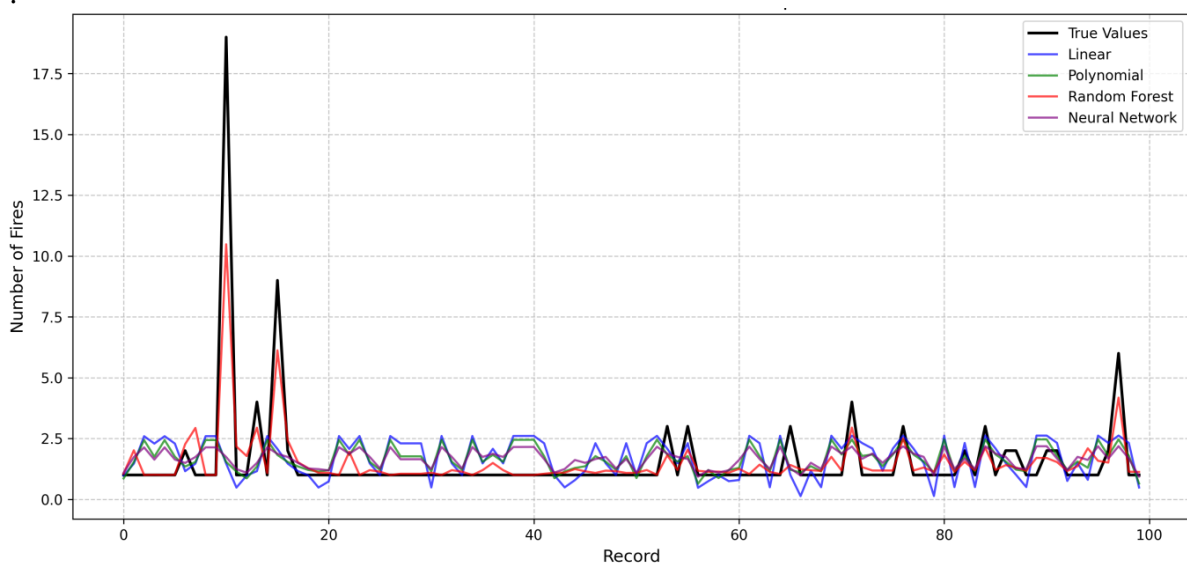


Рисунок 4 – Порівняння результатів прогнозування із реальними даними

На рисунку 5 наведено візуалізацію результатів прогнозування річної кількості пожеж для території Львівського району за допомогою

моделі випадковий ліс. Цифрами зазначено оціночну кількість пожеж за рік для кожного із шестикутників площею 3 км².

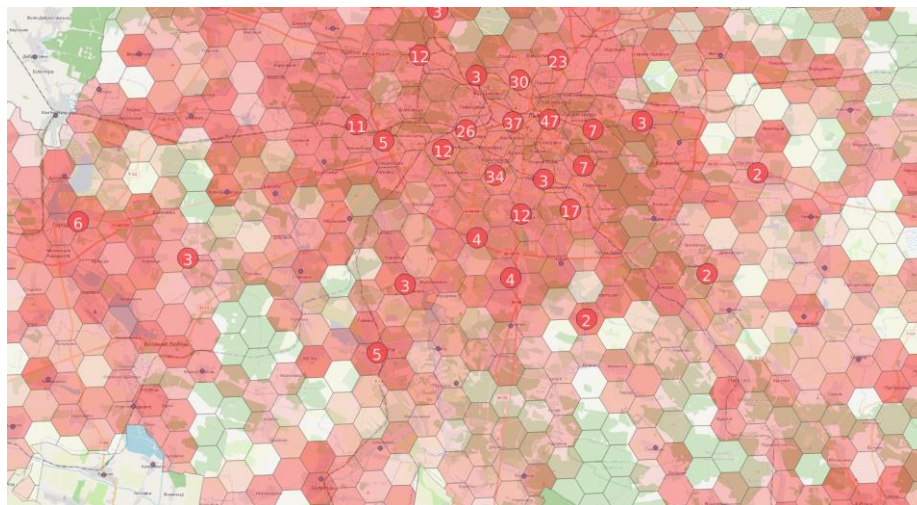


Рисунок 5 – Прогнозування кількості пожеж на території Львівського району

Розроблений метод дає змогу здійснювати оптимізацію районів виїздів пожежно-рятувальних підрозділів та розподілу матеріальних ресурсів. Окрім того, розроблений метод може бути використано для прогнозування та розрахунку ризиків виникнення інших надзвичайних подій, таких як дорожньо-транспортні пригоди, нещасні випадки на виробництві, пожежі в екосистемах тощо.

Висновки. Аналіз даних про виїзди пожежно-рятувальних підрозділів Львівської області показав, що ретельна підготовка даних і вибір моделі машинного навчання є ключовими для досягнення точних результатів прогнозування кількості пожеж у житловому секторі. Підготовка даних охоплювала нормалізацію даних про виклики, агрегацію інформації за шестикутниками площею 3 км² на основі даних за 2016-2021 роки, виділення пожеж у житловому секторі, включення предикторів густоти населення та місяців як *dummy*-змінних. Оцінка чотирьох моделей — лінійної регресії (MAE = 1,486, RMSE = 2,412, R^2 = 0,412), поліноміальної регресії (MAE = 1,412, RMSE = 2,342, R^2 = 0,445), випадкового лісу (MAE = 0,855, RMSE = 1,253, R^2 = 0,841) і нейронної мережі (MAE = 1,393, RMSE = 2,334, R^2 = 0,450) — виявила, що випадковий ліс значно перевершує інші методи за точністю й узагальнювальною силою завдяки своїй здатності вловлювати нелінійні залежності між густотою населення, місяцями та кількістю пожеж, тоді як лінійна й поліноміальна регресії показали слабкі результати через обмежену складність, а нейронна мережа – через недостатню оптимізацію й обсяг даних.

Запропонований метод прогнозування (на основі випадкового лісу) демонструє практичну

цінність для ДСНС України у Львівській області та інших органів влади, дозволяючи точно прогнозувати кількість пожеж у різних локаціях. Це створює передумови для оптимізації розподілу ресурсів і планування превентивних заходів, а його гнучкість дає змогу адаптувати модель до прогнозування інших надзвичайних ситуацій, таких як повені, дорожньо-транспортні пригоди чи техногенні аварії, за умови введення відповідних предикторів, що підвищує ефективність реагування й зменшує потенційні збитки. Подальші дослідження мають зосередитись на додаванні метеорологічних чи соціально-економічних предикторів, які дадуть змогу виконувати прогнозування ризиків в умовах реального часу, оптимізації параметрів випадкового лісу та порівнянні з іншими методами (наприклад, градієнтний бустинг) з метою подальшого підвищення точності прогнозування.

Список літератури:

1. Andreieva, O., Skydan, O., Wójcik, R., Kędziora, W., & Alpatova, O. (2022). Influence of weather conditions on the spread of fires in the forest fund of Zhytomyr Polissia. *Scientific Horizons*, 25(3), 68-75.
2. Zhisheng Xu, Dingli Liu, Long Yan, Temperature-based fire frequency analysis using machine learning: A case of Changsha, China, *Climate Risk Management*, Volume 31, 2021, <https://doi.org/10.1016/j.crm.2021.100276>.
3. Yang, L., Dawson, C. W., Brown, M. R., & Gell, M. (2006). Neural network and GA approaches for dwelling fire occurrence prediction. *Knowledge-Based Systems*, 19(4), 213–219. doi.org/10.1016/j.knosys.2005.11.021

4. Pahlavani, P., Raei, A., Bigdeli, B., & Ghorbanzadeh, O. (2024). Identifying Influential Spatial Drivers of Forest Fires through Geographically and Temporally Weighted Regression Coupled with a Continuous Invasive Weed Optimization Algorithm. *Fire*, 7(1), 33. <https://doi.org/10.3390/fire7010033>

5. Guan, Ruide. (2023). Predicting Forest Fire with Linear Regression and Random Forest. *Highlights in Science, Engineering and Technology*. 44. 1-7. [10.54097/hset.v44i.7159](https://doi.org/10.54097/hset.v44i.7159).

6. Bera, B., Shit, P. K., Sengupta, N., Saha, S., & Bhattacharjee, S. (2022). Forest fire susceptibility prediction using machine learning models with resampling algorithms, northern part of eastern Ghat mountain range (India). *Geocarto International*, 37(26), 11756–11781.

7. Damaševičius, R., Maskeliūnas, R. (2024). Distributed Random Forest for Predicting Forest Wildfires Based on Weather Data. In: Verma, A., Verma, P., Pattanaik, K.K., Dhurandher, S.K., Woungang, I. (eds) *Advanced Network Technologies and Intelligent Computing. ANTIC 2023. Communications in Computer and Information Science*, vol 2091. Springer, Cham. https://doi.org/10.1007/978-3-031-64064-3_22.

8. Safi, Youssef & Bouroumi, Abdelaziz. (2013). Prediction of forest fires using Artificial neural networks. *Applied Mathematical Sciences*. 7. 271-286. [10.12988/ams.2013.13025](https://doi.org/10.12988/ams.2013.13025).

9. Cortez, P., & Morais, A. (2007). A data mining approach to predict forest fires using meteorological data. In J. Neves, M. F. Santos, & J. Machado (Eds.), *New trends in artificial intelligence: Proceedings of the 13th EPIA 2007 Portuguese Conference on Artificial Intelligence* (pp. 512–523). Guimarães, Portugal.

10. Safi, Y., Bouroumi, A., & others. (2011). A neural network approach for predicting forest fires. In *Proceedings of the Second International Conference on Multimedia Computing and Systems (ICMCS'11)*. Ouarzazate, Morocco.

11. R. Ratushny, A. Tryhuba, O. Bashynsky and V. Ptashnyk, Development and Usage of a Computer Model of Evaluating the Scenarios of Projects for the Creation of Fire Fighting Systems of Rural Communities, 2019 XIth International Scientific and Practical Conference on Electronics and Information Technologies (ELIT), Lviv, Ukraine, 2019, pp. 34-39, doi: [10.1109/ELIT.2019.8892320](https://doi.org/10.1109/ELIT.2019.8892320).

12. Шопський О. М., Малець І. О. Аналіз і вдосконалення моделі кластеризації даних для формування вибірки з метою прогнозування ризикових ситуацій. Інформаційна безпека та інформаційні технології ІБІТ, Збірник доповідей V Міжнародної науково-практичної конференції. Львів, ЛДУ БЖД, 2024. С.379-381.

13. Великий електронний словник української мови (BECYM). URL: github.com/brown-uk/dict_uk

14. Chaitanya, B. S. S. K., Reddy, D. A. K., Chandra, B. P. S. E., Krishna, A. B., & Menon, R. R. K. (2019). Full-text search using database index. In *2019 5th International Conference on Computing, Communication, Control and Automation (ICCUBEA)* (pp. 1–5). IEEE. <https://doi.org/10.1109/ICCUBEA47591.2019.9128683>

15. Ю. О. Олійник, Система аналізу текстових потоків даних, ППММ, вип. 1, 2023, с. 149–158, doi: [10.32782/2618-0340/2020.1-3.15](https://doi.org/10.32782/2618-0340/2020.1-3.15).

References:

1. Andreieva, O., Skydan, O., Wójcik, R., Kędziora, W., & Alpatova, O. (2022). Influence of weather conditions on the spread of fires in the forest fund of Zhytomyr Polissia. *Scientific Horizons*, 25(3), 68–75.

2. Xu, Z., & Liu, D., & Yan, L. (2021). Temperature-based fire frequency analysis using machine learning: A case of Changsha, China. *Climate Risk Management*, 31, Article 100276. <https://doi.org/10.1016/j.crm.2021.100276>

3. Yang, L., Dawson, C. W., Brown, M. R., & Gell, M. (2006). Neural network and GA approaches for dwelling fire occurrence prediction. *Knowledge-Based Systems*, 19(4), 213–219. doi.org/10.1016/j.knosys.2005.11.021

4. Pahlavani, P., Raei, A., Bigdeli, B., & Ghorbanzadeh, O. (2024). Identifying influential spatial drivers of forest fires through geographically and temporally weighted regression coupled with a continuous invasive weed optimization algorithm. *Fire*, 7(1), Article 33. doi.org/10.3390/fire7010033

5. Guan, R. (2023). Predicting forest fire with linear regression and random forest. *Highlights in Science, Engineering and Technology*, 44, 1–7. <https://doi.org/10.54097/hset.v44i.7159>

6. Bera, B., Shit, P. K., Sengupta, N., Saha, S., & Bhattacharjee, S. (2022). Forest fire susceptibility prediction using machine learning models with resampling algorithms, northern part of eastern Ghat mountain range (India). *Geocarto International*, 37(26), 11756–11781.

7. Damaševičius, R., & Maskeliūnas, R. (2024). Distributed random forest for predicting forest wildfires based on weather data. In A. Verma, P. Verma, K. K. Pattanaik, S. K. Dhurandher, & I. Woungang (Eds.), *Advanced network technologies and intelligent computing: ANTIC 2023* (pp. 265–280). Springer. https://doi.org/10.1007/978-3-031-64064-3_22

8. Safi, Y., & Bouroumi, A. (2013). Prediction of forest fires using artificial neural networks. *Applied Mathematical Sciences*, 7, 271–286. <https://doi.org/10.12988/ams.2013.13025>

9. Cortez, P., & Morais, A. (2007). A data mining approach to predict forest fires using meteorological data. In J. Neves, M. F. Santos, & J. Machado (Eds.), *New trends in artificial intelligence: Proceedings of the 13th EPIA 2007 Portuguese Conference on Artificial Intelligence* (pp. 512–523). Guimarães, Portugal.
10. Safi, Y., Bouroumi, A., & others. (2011). A neural network approach for predicting forest fires. In *Proceedings of the Second International Conference on Multimedia Computing and Systems (ICMCS'11)*. Ouarzazate, Morocco.
11. R. Ratushny, A. Tryhuba, O. Bashynsky and V. Ptashnyk, "Development and Usage of a Computer Model of Evaluating the Scenarios of Projects for the Creation of Fire Fighting Systems of Rural Communities," 2019 XIth International Scientific and Practical Conference on Electronics and Information Technologies (ELIT), Lviv, Ukraine, 2019, pp. 34-39, doi: 10.1109/ELIT.2019.8892320.
12. Shopsyki, O. M., & Malets, I. O. (2024). Analysis and improvement of the data clustering model for sampling for the purpose of predicting risk situations. In *Information security and information technologies IBIT-2024: Collection of reports of the V International Scientific and Practical Conference* (pp. 379–381). Lviv, Ukraine: LSULS.
13. *Big Electronic Dictionary of the Ukrainian Language (VESUM)*. (n.d.). GitHub. github.com/brown-uk/dict_uk
14. Chaitanya, B. S. S. K., Reddy, D. A. K., Chandra, B. P. S. E., Krishna, A. B., & Menon, R. R. K. (2019). Full-text search using database index. In *2019 5th International Conference on Computing, Communication, Control and Automation (ICCUBEA)* (pp. 1–5). IEEE. doi.org/10.1109/ICCUBEA47591.2019.9128683
15. Oliynyk, Y. O. (2023). System for analyzing textual data streams. *PPMM*, 1, 149–158. <https://doi.org/10.32782/2618-0340/2020.1-3.15>

© О. М. Шопський, О. В. Хлевной, 2025.
Науково-методична стаття.
 Надійшла до редакції 22.04.2025.
 Прийнято до публікації 04.06.2025.