

Данило Вороновський, Ігор Пірко*

магістрант кафедри КН, НЛТУ України, Львів, Україна.

* доцент кафедри комп'ютерних наук, НЛТУ України, Львів, Україна.

Розроблення інформаційної системи аналізу передвісників землетрусів

Анотація – Землетрус це раптове трясіння поверхні землі, що виникає внаслідок раптового вивільнення енергії. Надійне передбачення землетрусів має численні соціальні та інженерні переваги. Останнім часом експоненційно зростаючий об'єм сейсмічних даних призвів до розробки кількох алгоритмів автоматичного виявлення землетрусів за допомогою підходів машинного навчання. В даній роботі були використані різні алгоритми для виявлення землетрусів, такі як дерево рішень, класифікатор випадкового лісу, машина опорних векторів (Support Vector Machine), класифікатор К-найближчих сусідів (KNeighborsClassifier).

Ключові слова – машинне навчання, передвісники землетрусів, Python, Matplotlib.

Поняття передвісника використовується в різних областях знань і означає явище, яке передуює деякій події і обов'язково пов'язане з цією подією. Підготовка землетрусу викликає збурення, які виділяються на фоні випадкових варіацій, різноманітних геофізичних полів, що трактуються як передвісники землетрусу. Передвісниками землетрусу називають зміни геофізичних полів, які спричинені процесом підготовки землетрусу [3].

Інформаційно-аналітична система прогнозу землетрусів розробляється з метою узагальнити дані по передвісниках поштовхів землі та скоректувати регіональні коефіцієнти закономірностей прояву передвісників. Землетруси являються одним з найбільш небезпечних природних явищ, в зв'язку з цим теорії виникнення передвісників, які використовуються для прогнозу землетрусів, надають великого значення. Землетруси повторюються в одних і тих же сейсмоактивних районах, але, очевидно, не підлягають якійсь закономірності. Накопичення енергії та напруженого стану земної кори, які фіксуються вздовж активних розломів, дають основу прогнозувати в цьому місці землетрус. Однак для

того, щоб передбачити час поштовхів та епіцентр землетрусу, потрібно обробити велику кількість незалежних даних.

Основою аналізу передвісників землетрусів являються кореляційні залежності між часом прояву передвісників та енергією землетрусу. В роботі вивчався вміст радону в ґрунтових водах. Передбачається, що передвісники виникають одночасно на різних відстанях від епіцентру майбутнього землетрусу. Але процес підготовки землетрусу відбувається не тільки в часі, але й в просторі, тому тривалість передвісників залежить від відстані пункту спостереження до епіцентру землетрусу.

Всі короткотривалі передвісники – це свідчення певного фізичного процесу в зоні можливого землетрусу, незалежно від методу спостережень вони повинні проявлятися в один і той же момент часу. Тривалість короткотермінових передвісників складає декілька годин. Тривалість процесу накопичення енергії, що являється причиною появи передвісників, збільшується з ростом магнітуди землетрусу.

Розроблено інформаційно-аналітичну систему для аналізу передвісників землетрусів. Однією з важливих частин підготовки даних був вибір конкретних ознак для подальшого аналізу та моделювання. Після цього використано бібліотеку Scikit-Learn для розділення даних на навчальний та тестовий набори. Для забезпечення стабільності та найкращої ефективності моделі використано StandardScaler для масштабування ознак. Особливу увагу приділено балансу класів, для цього було використано метод Smote для збільшення зразків меншого класу, чим вдалося вирівняти розподіл класів у навчальному наборі даних. Це важливий крок для покращення ефективності моделювання та уникнення перекосу у використанні даних [1]. Було використано бібліотеку Seaborn для побудови графіків та гістограм, зокрема, для візуалізації розподілу деяких ознак у наборі даних, що допомагає краще зрозуміти характеристики даних.

	title	magnitude	date_time	cdi	mmi	alert	Earthquake	sig	net	nst	dmin	gap	magType	depth	:
0	M 7.0 - 18 km SW of Malango, Solomon Islands	7.0	22-11- 2022 02:03	8	7	green		1	768	us	117	0.509	17.0	mww	14.000
1	M 6.9 - 204 km SW of Bengkulu, Indonesia	6.9	18-11- 2022 13:37	4	4	green		0	735	us	99	2.229	34.0	mww	25.000
2	M 7.0 -	7.0	12-11- 2022 07:09	3	3	green		1	755	us	147	3.125	18.0	mww	579.000
	M 7.3 - 205 km		11 11												

Рис. 1. Структура датасету для подальшого аналізу.

ММІ (змінена інтенсивність Меркаллі) враховує низку факторів, таких як сила поштовху, пошкодження будівель та інфраструктури тощо. sig: скорочення від magnitude significance, це міра відносної сили землетрусу. Розраховується на основі магнітуди землетрусу, кількості сейсмічних станцій, які повідомляють про подію, і відстань між подією та цими станціями. net: скорочення від seismic network, це відноситься до мережі сейсмометрів та інших сейсмічних датчиків, що використовуються для виявлення та реєстрації сейсмічних подій. Кожна сейсмічна мережа має свої специфічні характеристики та можливості, які можуть вплинути на точність і повноту виявлення землетрусу та місцезнаходження. nst: скорочення від number of seismic stations, це кількість сейсмічних станцій, які виявили та зафіксували землетрус. Чим більше сейсмічних станцій виявляють подію, тим точніше її місцезнаходження та інше можна визначити його атрибути. dmin: скорочення від minimum distance, це найкоротша відстань між землетрусом і найближчою сейсмічною станцією, яка це виявила. Ця відстань може вплинути на точність визначення місця землетрусу та інші атрибути, а також силу сейсмічних хвиль, виявлених станцією. Азимутальний розрив – кут між двома найбільш віддаленими сейсмічними станціями, які фіксують землетрус. Великий розрив може вказувати на відсутність важливих сейсмічних даних, що може ускладнити точне визначення місця землетрусу, магнітуди та інших характеристик. magType відноситься до типу обчислення магнітуди,

який використовується для визначення розміру землетрусу [2]. Є кілька різних типів розрахунків величини, включаючи величину Ріхтера, величину моменту та поверхню величина хвилі. Кожен тип обчислення магнітуди враховує різні аспекти землетрусу, наприклад яка амплітуда сейсмічних хвиль або енергія, вивільнена під час землетрусу. Вибір типу величини може впливати на зареєстроване значення магнітуди, а також можуть впливати на оцінку небезпеки землетрусу та планування реагування на надзвичайні ситуації.

Було підготовлено дані для подальшого моделювання та аналізу. Ці кроки стали основою для розробки та оцінки моделі машинного навчання для аналізу передвісників землетрусів та пов'язаних з ними сигналів [3]. На основі набору даних, який містить інформацію про землетруси, включаючи різні ознаки, такі як довгота, широта, глибина, cdi, mmi, sig, dmin, nst, gap і колір. Було використано бібліотеку Seaborn для візуалізації даних, зокрема для побудови гістограм ознак магнітуди – MMI. Для візуалізації графіків також використано бібліотеку Matplotlib, створюючи підграфіки і додавши заголовки та підписи до осей.

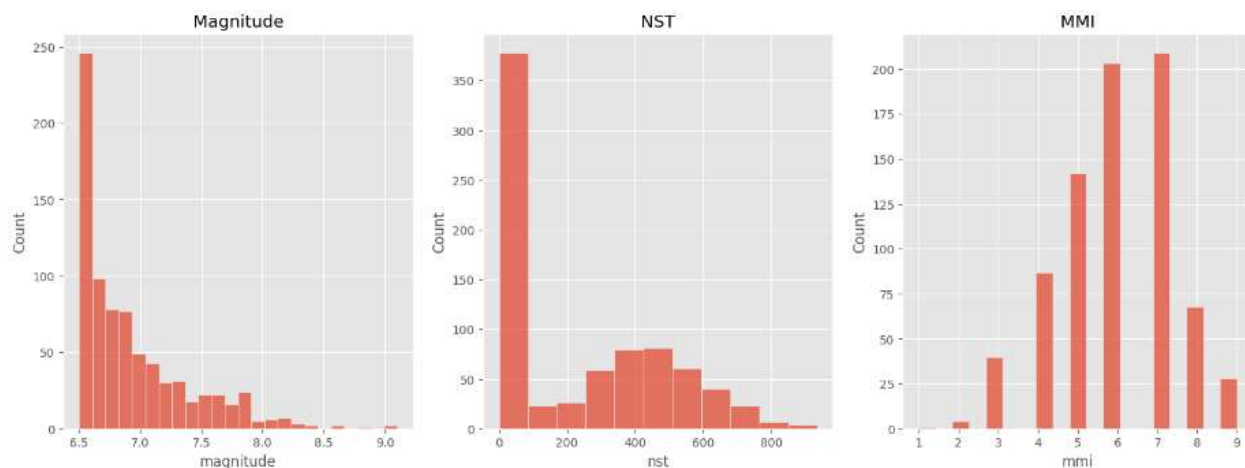


Рис. 2. Найголовніші вхідні ознаки моделі.

Було проведено навчання кількох класифікаторів машинного навчання, таких як DecisionTreeClassifier, KNeighborsClassifier, RandomForestClassifier, на вхідних даних. Після навчання моделей їх було збережено у списку models. Обчислювалися різні метрики, такі як accuracy, precision, recall і F1-оцінка для оцінки ефективності кожної моделі. Для порівняння результатів

було створено таблицю, в якій відображені значення цих метрик для різних моделей:

	Name	Accuracy	Precision_score	Recall_score	F1_score
0	DecisionTreeClassifier	0.942308	0.941710	0.942308	0.941875
1	KNeighborsClassifier	0.923077	0.934314	0.923077	0.923851
2	SVC	0.930769	0.945339	0.930769	0.932321
3	RandomForestClassifier	0.976923	0.977345	0.976923	0.976927

Для оцінки взаємозв'язку вхідних ознак моделі створено теплову карту, яка графічно відображає кореляцію між числовими змінними. З неї можна зробити висновок про те, які змінні корелюють між собою та в якому напрямку (позитивна або негативна кореляція). Матриця кореляції показує ступінь взаємозв'язку між парами змінних. Значення кореляції від -1 до 1, де -1 вказує на повну негативну кореляцію, 1 – на повну позитивну кореляцію, а 0 – на відсутність кореляції [2]. З допомогою параметра `annot=True` вкажуть, щоб значення кореляції були виведені на самій тепловій карті. Ці значення будуть розміщені в кожній клітинці теплової карти, що відповідає парі змінних.

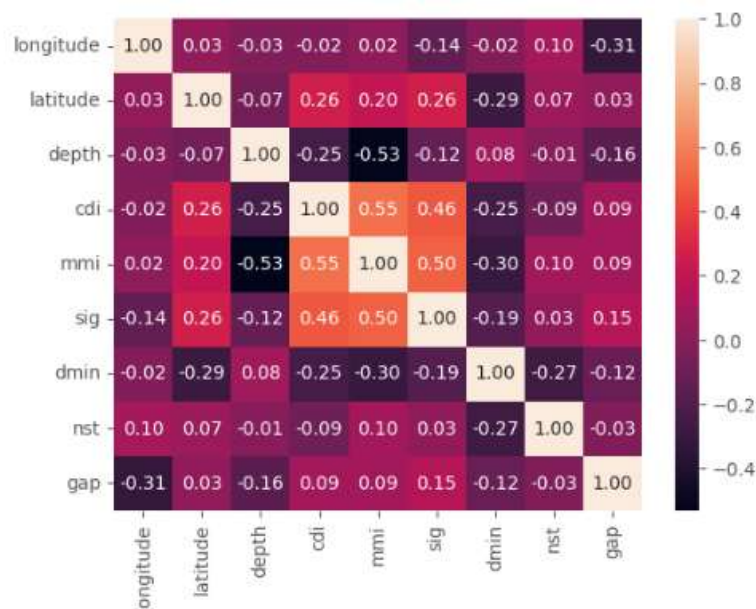


Рис. 3. Теплова карта кореляції між вхідними ознаками.

Ці моделі було використано для передбачення категорій магнітуди землетрусу на основі введених користувачем даних. Категорії магнітуди землетрусу визначаються за кольорами, де зелений вказує на низьку магнітуду, жовтий – середню, помаранчевий – високу і червоний – дуже високу магнітуду.

Висновки. Розроблено програмне забезпечення сейсмологічної аналітичної системи, яка розділена на дві основні частини: підсистему вводу, зберігання та управління даними та підсистему аналізу даних. Підсистема аналізу даних реалізована з використанням різних бібліотек машинного навчання для вирішення конкретних завдань у контексті системи.

Список використаних літературних джерел

- [1] Элбон Крис. Машинное обучение с использованием Python. Сборник рецептов / – СПб: БХВ-Петербург, 2019. – 384 с.
- [2] Шамин Р.В. Практическое руководство по методам машинного обучения / – М.: Lector.ru, 2019. – 93 с.
- [3] Завьялов А.Д. Среднесрочный прогноз землетрясений. Основы, методика, реализация / – М.: Наука, 2006. – 256 с.

Oleh Haba

assistant teacher of Computer Science Department, Ukrainian National Forestry University, Lviv, Ukraine, <https://orcid.org/0009-0006-2140-4524>.
oleh.haba@nltu.edu.ua

The use of artificial intelligence in database and knowledge sytems - an analysis of machine learning, neural networks, and other ai methods for database and knowledge analysis

Abstract - This presentation explores the use of artificial intelligence (AI) in database and knowledge systems, with a focus on the applications of machine learning, neural networks, and other AI methods for data and knowledge analysis. Machine learning allows computers to learn from data and make predictions, while neural networks excel in natural language processing. Processing large volumes of data with AI simplifies analysis and provides personalized recommendations. Furthermore, AI aids in automating the analysis of natural data, such as medical images. The presentation concludes by emphasizing the profound impact of AI on contemporary data management and knowledge utilization.