

У. Ю. Дзелендзяк¹, В. В. Самотий^{2,3}, О. М. Пелех¹

¹Національний університет “Львівська політехніка”,
кафедра комп’ютеризованих систем автоматики

²Львівський державний університет безпеки життєдіяльності,
кафедра управління інформаційною безпекою

³Краківська політехніка ім. Тадеуша Костюшка,
кафедра автоматики та інформаційних технологій

РОЗРОБЛЕННЯ МОДЕЛІ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ПРОГНОЗУВАННЯ ДОРОЖНЬО-ТРАНСПОРТНИХ ПРИГОД

© Дзелендзяк У. Ю., Самотий В. В., Пелех О. М., 2018

Розроблено систему для аналізу чинників, що впливають на виникнення дорожньо-транспортних пригод (ДТП), за допомогою інструментів та бібліотек мови Python. Використовуючи досліджені чинники, розроблено та досліджено нейронну мережу для прогнозування можливості виникнення ДТП. На основі спроектованої нейронної мережі розроблено веб-платформу для автоматичного прогнозування можливості виникнення ДТП та візуального відображення обчислених ризиків на найближчу добу.

Ключові слова: штучна нейронна мережа, регресійна модель, багат шаровий перцептрон, веб-платформа, дорожньо-транспортні пригоди.

The system for factor analysis, which makes influence on road accidents appearance has been developed using Python language tools and libraries. Using investigated factors the neural network for possibility of road accident forecasting has been developed and created. Based on designed neural network the web platform for automatic forecasting of road accidents possibility appearance and visual reflection of calculated risks for the next day has been created.

Key words: artificial neural network, regression model, multilayer perceptron, web platform, road accidents.

Вступ

Дорожньо-транспортні пригоди є гострою проблемою, однаковою мірою і у країнах, що розвиваються, і у розвинених країнах, адже це негативний чинник для розвитку найважливіших сфер: охорони здоров’я, соціальних, екологічних та економічних. Проаналізувавши дані аварійності на конкретній території за певний період, можна спрогнозувати ймовірність для автотранспорту потрапити у ДТП. Тому розроблення системи для аналізу та прогнозування виникнення ДТП є актуальною проблемою, оскільки сьогодні в Україні не існує системи для прогнозування виникнення ДТП.

На території міста Львова протягом останніх років спостерігається значне зростання кількості учасників дорожнього руху, тому аналіз розподілу дорожньо-транспортних пригод, а також візуальне відображення та прогнозування дорожньо-транспортних пригод з метою запобігання та мінімізації трагічних випадків та завдання шкоди здоров’ю і майну учасників дорожнього руху набуває все більшого значення.

Аналіз публікацій

Згідно з офіційною статистикою [6], щороку в дорожньо-транспортних пригодах (ДТП) в Україні травмуються більше ніж 30 тис. осіб та понад 3 тис. осіб гине. Згідно з даними Опендатабот [4] за січень-квітень 2018 року Львівська область посідає друге місце в Україні за кількістю ДТП і перше місце за кількістю смертельних ДТП. Сьогодні система безпеки транспортного руху в Україні спирається на засоби профілактики і попередження ДТП, хоча сучасний розвиток інформаційних технологій та наукових методів дає змогу діяти на випередження – прогнозувати, а відповідно, і запобігати виникненню аварійних ситуацій на транспортних шляхах, цілеспрямовано здійснюючи вплив там, де аварія ще не відбулася, але висока її ймовірність. Хоча конкретну аварію практично неможливо передбачити, через випадкову природу ДТП, дослідники визначили, що агрегація великої кількості аварій на достатню площу території та інтервал часу дає можливість встановити певний рівень передбачуваності, який можна описати математичними та статистичними зв'язками, беручи до уваги різноманітні фактори впливу (інтенсивність трафіку, погодні умови) [9]. На жаль, в Україні немає централізованих детальних даних про ДТП, які давали б достатньо інформації для аналізу причинно-наслідкових подій аварійності на автошляхах країни.

Розроблення системи для аналізу дорожньо-транспортних пригод

Проаналізуємо інформацію про зареєстровані у 2015 р. ДТП у місті Львові, опубліковану на Порталі відкритих даних Львова [2] у CSV-файлі, який містить дані про тип, дату, час скоєння та адресу ДТП. Як інструмент для аналізу та візуалізації великих наборів даних вибрано Python, який зарекомендував себе як універсальний інструмент для різноманітних наукових завдань. Для оброблення та підготування даних використано Python-інструмент Pandas, із допомогою якого дані зчитано із CSV-файла, відформатовано, конвертовано у необхідні типи та проаналізовано частково відсутні дані. Наступний крок – конвертування адрес ДТП у формат просторових даних та прив'язування географічних координат до кожної адреси. Для цього завдання використано Python-бібліотеку GeoPy, що дає змогу проводити геокодування за допомогою систем геокодерів. Як систему геокодування вибрано картографічний сервіс HERE. Для геокодування у бібліотеці GeoPy є функція geocode, що приймає як параметр адресу, яку потрібно конвертувати у просторові дані.

Для ефективного аналізу погодних факторів впливу потрібно дізнатися метеорологічні умови для кожного окремого випадку ДТП. Такі дані отримано за допомогою сервісу Dark Sky, який надає доступ до історичних погодних даних будь-якої точки світу. Щоб взаємодіяти із Dark Sky, потрібно зареєструватися та отримати унікальний API ключ. Для зручності одержання даних Dark Sky пропонує бібліотеки-обгортки. Для роботи вибрано бібліотеку-обгортку Dark Sky API для Python 3 – Forecastiory. Ця бібліотека містить дев'ять класів, основний з яких – ForecastIO здійснює підключення, формує URL-адресу запиту та отримує дані з Dark Sky API. Для того, щоб дізнатися, у світлу чи у темну пору доби трапилася ДТП, використано Python-бібліотеку ephem, яка надає інструменти для виконання астрономічних обчислень.

Найбільша концентрація аварій із постраждалими спостерігається на вулицях у центрі міста: проспекті Свободи та вул. П. Дорошенка (рис. 1), 63 % аварій, у яких є постраждалі, трапилися в світлу пору доби. Це пов'язано із тим, що трафік у денний час насиченіший. Як свідчать дані досліджень [11], [13] та статистики [12], максимальна кількість ДТП, у яких постраждали чи загинули люди, спостерігається у суботу та неділю, а мінімальна – у вівторок. Аналіз даних аварійності міста Львова має дві відмінності: пік аварій із постраждалими зміщується із суботи–неділі, як вказано у згаданих вище роботах, на п'ятницю–суботу; мінімальна кількість аварій із постраждалими у Львові спостерігається у неділю, тоді як у згаданих роботах – у вівторок (у Львові на вівторок припадає мінімум протягом робочого тижня). На рис. 2 наведено десять вулиць міста з найбільшою кількістю ДТП, у яких є постраждалі за даними 2018 р.

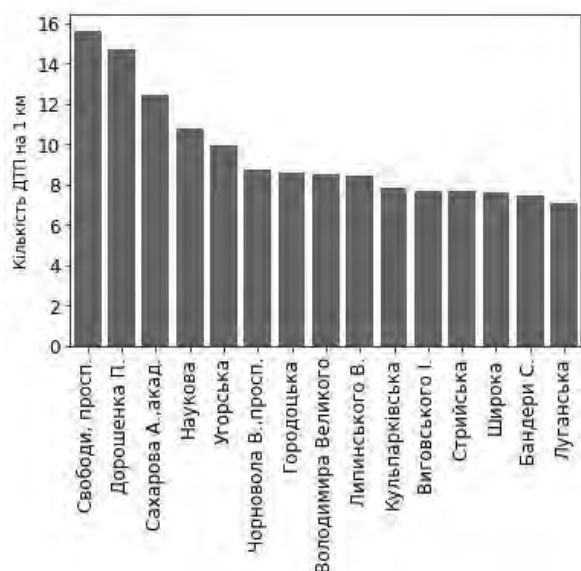


Рис. 1. Кількість ДТП із постраждалими на 1 км

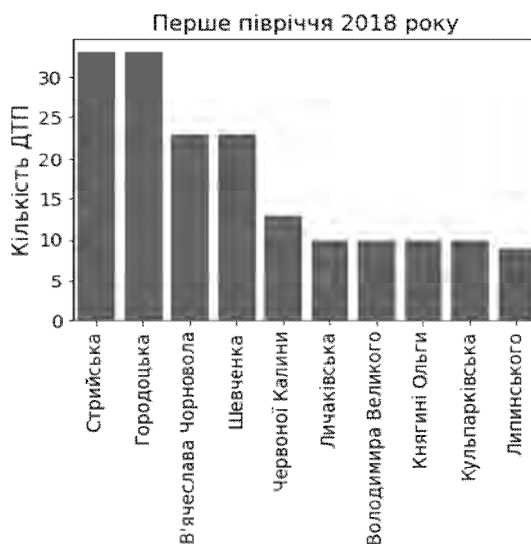


Рис. 2. Топ-10 вулиць міста за кількістю ДТП, у яких є постраждалі

Модель нейронної мережі для прогнозування дорожньо-транспортних пригод

Прогнозування можливості виникнення ДТП із постраждалими на території міста Львова на основі встановлених показників, які впливають на їх виникнення, здійснювалося за допомогою штучної нейронної мережі (ШНМ) за регресійною моделлю. Вхідні параметри ШНМ наведено у табл. 1, а її схему на рис. 3.

Таблиця 1

Вхідні показники ШНМ

Параметр	Позначення	Тип
Регіон	X_p	Елемент зі списку районів міста
Час дня	$X_{чд}$	Елемент зі списку (ранок, день, вечір, ніч)
День тижня	$X_{дт}$	Елемент зі списку днів тижня
Місяць року	$X_{мр}$	Елемент зі списку місяців у році
Пора доби	$X_{пд}$	Елемент зі списку (темна/світла пора доби)
Температура	X_t	Кількісна величина, °C
Відносна вологість	$X_{вв}$	Кількісна величина, %
Атмосферний тиск	$X_{ат}$	Кількісна величина, мм рт. ст
Швидкість вітру	$X_{шв}$	Кількісна величина, м/с
Хмарний покрив неба	$X_{хп}$	Кількісна величина (відсоток неба, закритого хмарами)
Видимість	X_v	Кількісна величина, км

Для вирішення завдань прогнозування найчастіше використовують нейронні мережі на основі багатозарового перцептрона (БШП).

Категоріальні дані (регіон, час дня, день тижня, місяць року, пора доби) перед подаванням на вхід нейронній мережі закодовано унітарним кодом. Основне завдання такого кодування – представлення ознаки категорії у векторному просторі із розмірністю, яка дорівнює кількості наявних категорій. Для цього встановлюють значення координати цієї категорії 1, а решти координат – 0.

Навчання, тестування та перевірку мережі здійснено на трьох вибірках випадково вибраних даних у процентному співвідношенні 70:20:10 відповідно. Проаналізовано 200 нейромереж, серед яких вибрано 5 із кращими результатами (табл. 2).

Таблиця 2

Характеристики підібраних архітектур ШНМ на основі БШП для регресійного аналізу умов виникнення ДТП із постраждалими у районах міста Львова

№	Архітектура нейромережі	Продуктивність навчання	Тестова продуктивність	Контрольна продуктивність	Похибка навчання	Тестова похибка	Контрольна похибка	Метод оптимізації	Функція похибки	Функція активації прихованих шарів	Функція активації вихідного шару
1	MLP 92-11-1	0,4038	0,3536	0,3056	0,0308	0,0357	0,0340	BFGS 28	SOS	Logistic	Identity
2	MLP 92-6-1	0,3947	0,3482	0,3033	0,0310	0,0357	0,0339	BFGS 27	SOS	Logistic	Identity
3	MLP 92-9-1	0,4221	0,3440	0,3039	0,0302	0,0358	0,0338	BFGS 30	SOS	Logistic	Identity
4	MLP 92-8-1	0,3932	0,3441	0,3067	0,0310	0,0358	0,0338	BFGS 25	SOS	Logistic	Identity
5	MLP 92-9-1	0,4019	0,3293	0,3041	0,0308	0,0362	0,0338	BFGS 27	SOS	Logistic	Identity

У вибраних БШП контрольна похибка лежить у межах від 0,0338 до 0,034 та продуктивність навчання від 0,3932 до 0,4221.

Для всіх вибраних БШП використано ітеративний метод оптимізації BFGS, який вважається одним із найефективніших, є стійким та має самокоригувальні властивості.

Для прихованих шарів нейронів функцією активації вибрано логістичну функцію, а для вихідного шару – лінійну, за допомогою якої рівень активації передається безпосередньо як вихід нейрона.

Для подальшого аналізу вибрано мережу із найменшим значенням похибки регресії – MLP 92-9-1 (рис. 4).

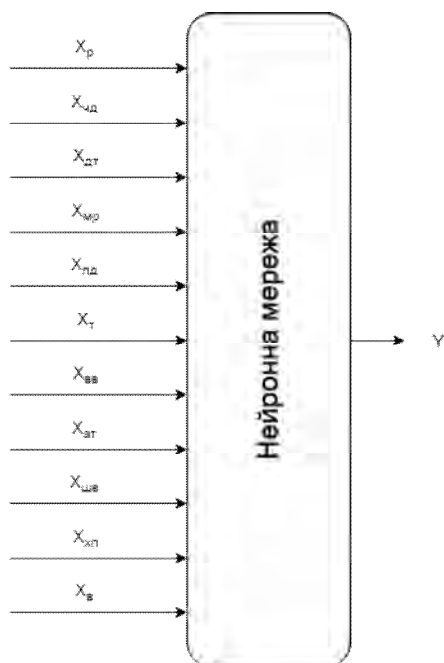


Рис. 3. Схема моделі ШНМ для регресійного аналізу умов виникнення ДТП із постраждалими у районах міста Львова

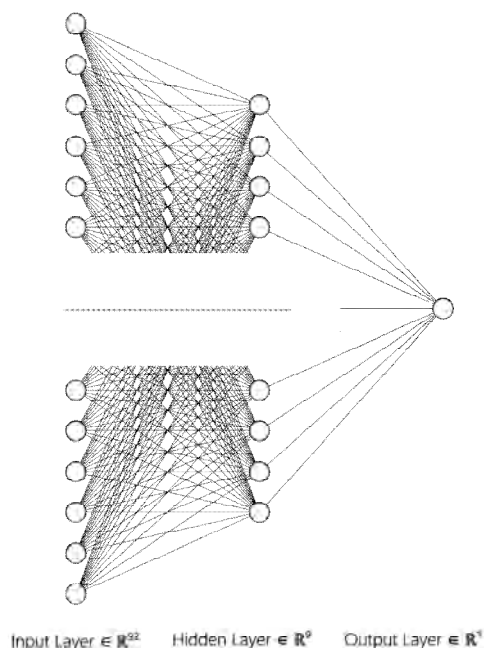


Рис. 4. Архітектура БШП 92-9-1 для регресійного аналізу умов виникнення ДТП із постраждалими у районах міста Львова

Аналіз результатів

Перевірка точності прогнозу мережі основана на подаванні на її входи 10 % від загальної кількості випадків, які не брали участі у навчанні мережі, вибраних довільно, частину з яких наведено у табл. 3.

Підвищення точності прогнозу можна досягти, взявши до уваги чинники, які можуть впливати на кількість ДТП, але вплив яких не досліджувався через їхню недоступність чи великий масштаб опрацювання, а саме густину населення у районах міста, стан дорожнього покриття, наявність факторів, що відволікають (наприклад, рекламних банерів).

Таблиця 3

Результат прогнозування ДТП із постраждалими у районах міста Львова розробленою ШНМ

№	Вхідні дані											Очікуваний результат	Результат на виході мережі
	temperature	humidity	pressure	windSpeed	cloudCover	visibility	region	daytime	month	day	is_daytime		
1	11	1,0	1014	0,0	0,6	10	16	3	7	5	0	0	0,079731
2	0	1,0	1025	3,0	1,0	0	18	2	11	4	0	0	-0,019942
3	2	0,7	1017	0,0	0,0	9	17	0	3	11	1	1	1,137844
4	3	0,9	1029	2,0	1,0	1	18	0	12	4	0	0	0,028706
5	3	0,9	1016	3,0	1,0	10	6	3	3	6	0	0	0,041600
6	12	0,9	1015	1,1	1,0	9	37	3	6	4	0	0	0,046039
7	13	1	1015	3	1	10	15	2	3	3	0	0	-0,146190
8	2	1,0	1020	1,0	1,0	5	30	3	11	1	0	0	0,051324
9	12	0,4	1015	3,8	0,6	9	6	2	3	24	1	1	0,873054
10	13	0,6	1018	4,4	0,8	10	7	1	9	1	1	1	0,739059
11	0	0,9	1010	6,0	0,8	10	31	2	1	7	0	0	0,281783
12	-1	0,7	1016	1,9	1,0	10	30	2	3	22	0	1	0,923681
13	-11	0,8	1019	3,0	0,8	10	2	2	1	2	0	0	0,250792
14	22	1	1014	6	0	10	2	2	6	1	1	0	0,240681
15	8	1	1024	2	1	4	28	1	11	6	1	0	0,221717
16	6	1	1019	2	1	0	28	3	10	2	0	0	0,072140
17	13	1	1018	3	1	5	7	3	9	6	0	0	0,070213
18	13	1	1014	0	1	10	0	3	7	5	0	0	0,032819
19	10	0	1012	6	1	10	13	2	3	5	1	0	-0,117559
20	9	0	1019	0	0	8	38	2	3	9	1	1	1,123866
21	9	1	1016	3	1	10	7	3	3	27	0	1	0,731271

Для відображення прогнозу можливості виникнення дорожньо-транспортних пригод із постраждалими у районах міста Львова створено веб-сайт (рис. 5). Цей веб-сайт в онлайн-режимі дає змогу дізнатись прогнозований розробленою нейронною мережею ступінь ймовірності виникнення аварій на дорогах для кожного району міста Львова у балах від 1 до 5, де 1 – низька ймовірність виникнення ДТП, 5 – ДТП, дуже ймовірно, трапиться у цьому районі в цей період часу.

З рис. 5 бачимо, що розроблена система аналізу та прогнозування виникнення ДТП дозволяє в режимі онлайн візуалізувати актуальний прогноз потенційних можливостей виникнення аварій із постраждалими на дорогах міста на найближчу добу.

Розроблена система є зручною та інтуїтивно зрозумілою для кінцевого користувача платформою для відображення ризиків виникнення ДТП із постраждалими. Її можуть використовувати як водії, для вибору оптимального маршруту переміщення містом, щоб уникнути потенційно небезпечних ділянок, так і представники правоохоронних органів, для здійснення дій на випередження виникнення потенційно небезпечних ситуацій.

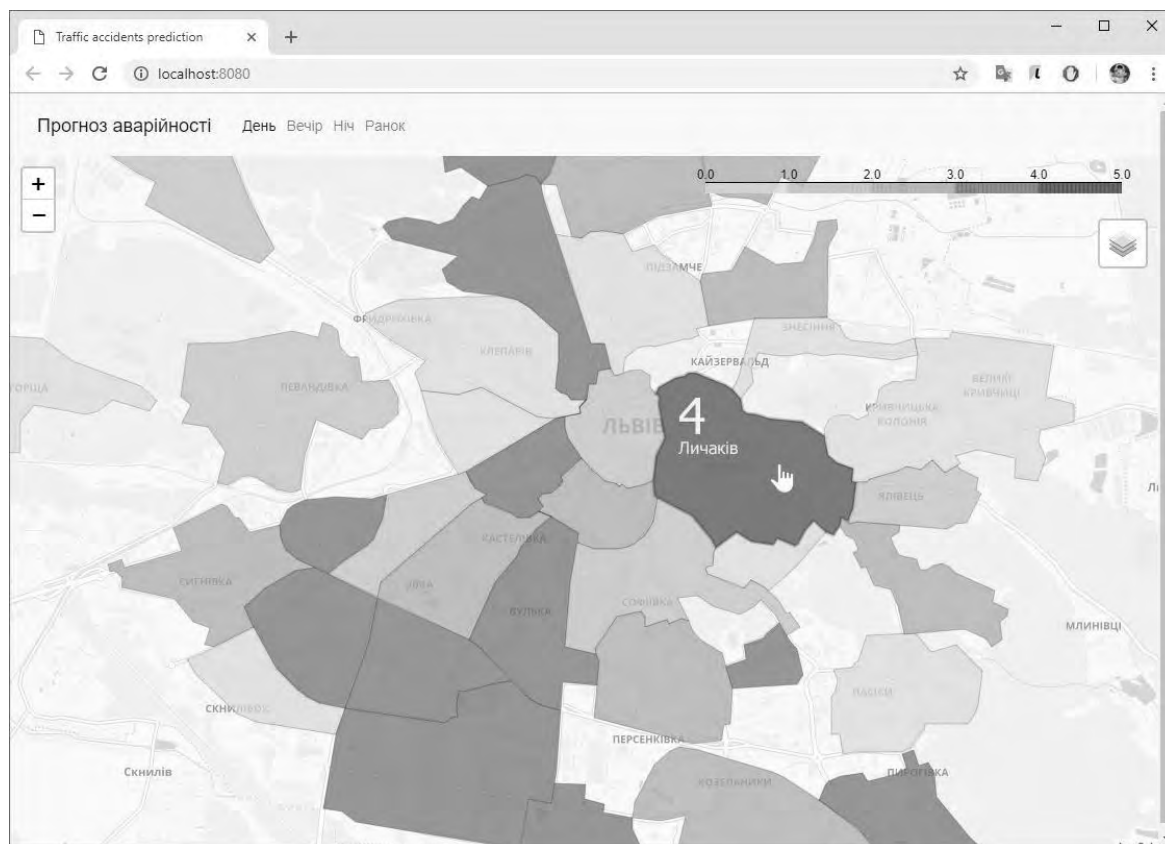


Рис. 5. Вигляд розробленого веб-сайта

Область використання цієї системи лежить у межах практичних застосувань для прогнозування ризиків виникнення нещасних випадків із постраждалими на дорогах міста, а також для наукових досліджень ефективності систем прогнозування на основі нейронних мереж.

Висновок

У статті наведено дослідження впливу погодних, часових та просторових умов на виникнення ДТП із постраждалими у межах міста Львова. Дослідження проведено у середовищі Jupyter Notebook за допомогою інструментів та бібліотек мови Python, а саме Pandas, NumPy, GeoPy, Matplotlib, Folium, PyEphem. Для аналізу використано дані ДТП, надані Національною поліцією України; погодні дані, отримані із платформи Dark Sky; дані меж районів міста, отримані із Платформи відкритих даних Львова, розпорядником яких є Інститут міста; дані геолокацій із сервісу HERE Maps із допомогою Python-бібліотеки GeoPy. Для візуалізації використано інструменти Matplotlib та Folium.

Досліджено чинники, що впливають на виникнення ДТП із постраждалими у межах міста Львова. Беручи до уваги ці чинники, розроблено та досліджено нейронну мережу для прогнозування можливості виникнення ДТП із постраждалими у районах міста.

На основі спроектованої нейронної мережі розроблено веб-платформу для автоматичного прогнозування можливості виникнення ДТП із постраждалими у районах міста Львова та візуального відображення обчислених ризиків на найближчу добу.

Розроблена система для аналізу впливу погодних, часових та просторових чинників на виникнення ДТП актуальна для удосконалення системи безпеки дорожнього руху та має практичне значення. Ця система універсальна, оскільки її можна використати для аналогічних досліджень для інших міст країни.

1. Адресний реєстр Львова – Відкриті дані Львова [Електронний ресурс] // ЛКП “Міський центр інформаційних технологій”. 2018. Режим доступу: <https://opendata.city-adm.lviv.ua/dataset/adreslviv>. 2. Інформація про ДТП 2015 [Електронний ресурс] // Портал відкритих даних Львова. – 2016. Режим доступу: <https://opendata.city-adm.lviv.ua/dataset/dtp2015>. 3. Мікрорайони Львова – Набори даних – Відкриті дані Львова [Електронний ресурс] // КУ Інститут міста. – 2018. – Режим доступу: <https://opendata.city-adm.lviv.ua/dataset/mikrorayony-lvova>. 4. Названі області, в яких частіше гинуть люди під час ДТП [Електронний ресурс] // Оpendatabot. – 2018. Режим доступу: <https://opendatabot.ua/blog/193-dtp-regions>. 5. Dark Sky API: Documentation Overview [Електронний ресурс] // The Dark Sky Company, LLC – Режим доступу: <https://darksky.net/dev/docs>. 6. Death on the roads [Електронний ресурс] // <http://roads.live.kiln.digital/>. – 2015. Режим доступу: World Health Organization. 7. Hernandez III. A Python wrapper for the forecast.io API [Електронний ресурс] / Hernandez III. – 2017. – Режим доступу: <https://github.com/bitpixedigital/forecastiopy3>. 8. How Do Weather Events Impact Roads? [Електронний ресурс] // Federal Highway Administration. – 2018. Режим доступу: https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm. 9. Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts / E. Hermans, T. Brijs, T. Stiers, C. Offermans. // Accident Analysis & Prevention. – 1995. – № 27. – С. 1–20. 10. Python Data Analysis Library [Електронний ресурс]. – 2018. – Режим доступу: <https://pandas.pydata.org/>. 11. Rural road safety – overview of crash statistics / M. Symmons, N. Haworth та I. Johnston // Monash University Accident Research Centre – 2014. 12. Statistics of fatal and injury road accidents in Lithuania, 2013–2016 / Lithuanian Road Administration – 2017. 13. Symmons M. Rural road safety - overview of crash statistics / M. Symmons, N. Rizenbergs, I. Johnston. – Melbourne: Monash University Accident Research Centre, 2014. 14. Welcome to GeoPy’s documentation! [Електронний ресурс] // GeoPy Contributors. – 2018. Режим доступу: <https://geopy.readthedocs.io/en/stable/>.