

Державна служба України з надзвичайних ситуацій
Львівський державний університет безпеки життєдіяльності
Навчально-науковий інститут цивільного захисту
Кафедра інформаційних технологій та систем електронних комунікацій

«Допущено до захисту»
Начальник кафедри ІТтаСЕК
підполковник служби цивільного
захисту
_____ Олександр ПРИДАТКО
“ _____ ” _____ 20__ року

ДИПЛОМНА РОБОТА МАГІСТРА

на тему «Інформаційна система інтелектуального аналізу даних дорожньо-транспортних пригод з використанням середовища *Jupyter*»

Виконав:
здобувач VI курсу, групи КН-61мз
спеціальності (освітньої програми)
122 «Комп'ютерні науки» (Комп'ютерні науки)
(шифр і назва спеціальності (освітньої програми))

_____ Олег БЛОКОНЬ

(ім'я та прізвище)

Керівник _____ Ольга СМОТР

(ім'я та прізвище)

Рецензент _____

(ім'я та прізвище)

Львів – 2023 року

Державна служба України з надзвичайних ситуацій
Львівський державний університет безпеки життєдіяльності
Навчально-науковий інститут цивільного захисту

Кафедра інформаційних технологій та систем електронних комунікацій

Освітній ступінь магістр

Спеціальність 122 “Комп’ютерні науки”

Освітня програма Комп’ютерні науки

ЗАТВЕРДЖУЮ

Начальник кафедри ІТтаСЕК
підполковник служби цивільного
захисту

Олександр ПРИДАТКО
“___” _____ 20__ року

ЗАВДАННЯ

на дипломну роботу

Здобувачу Олегу БІЛОКОНЮ

(ім’я, прізвище)

1. Тема «Інформаційна система інтелектуального аналізу даних дорожньо-транспортних пригод з використанням середовища Jupyter»

керівник роботи Ольга СМОТР, к.т.н., доцент

(ім’я, прізвище, науковий ступінь, вчене звання)

затверджені наказом ЛДУ БЖД від “__” _____ 202_ року №__ од

2. Термін подання здобувачем роботи _____ 2023 року

3. Початкові дані до роботи

1. Wes M. Python for Data Analysis / McKinney Wes
2. Kumar A. Python: Advanced Predictive Analytics / A. Kumar, J. Babcock., 2017. – 660 с. – (Packt).
3. Miller T. Modeling Techniques in Predictive Analytics with Python and R: A Guide to Data Science / Thomas Miller. – 448 с.
4. Xia, D., et al.: A distributed WND-LSTM model on MapReduce for short-term traffic flow prediction. Neural Comput. Appl. 33(7), 2393–2410 (2020). <https://doi.org/10.1007/s00521-020-05076-2>

5. “How Connected Vehicles Work | US Department of Transportation,” U.S. Department of Transportation, (2020). <https://www.transportation.gov/research-and-technology/how-connected-vehicles-work>, Accessed 12 May 2021

4. Зміст дипломної роботи/проекту (перелік питань, які потрібно розробити)

Вступ

Розділ 1. Предметна область. Збір інформації та аналіз.

Розділ 2. Дослідження методів та засобів реалізації інформаційної системи інтелектуального аналізу даних дорожньо-транспортних пригод

Розділ 3. Реалізація інформаційної системи інтелектуального аналізу даних дорожньо-транспортних пригод з використанням середовища Jupyter

Висновки

Список використаних літературних джерел

5. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

6. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів виконання дипломної роботи/проекту	Термін виконання етапів роботи	Примітка
1.	Предметна область. Збір інформації та аналіз		
2.	Дослідження методів та засобів реалізації інформаційної системи інтелектуального аналізу даних дорожньо-транспортних пригод		
3.	Реалізація інформаційної системи інтелектуального аналізу даних дорожньо-транспортних пригод з використанням середовища Jupyter		

Здобувач _____
(підпис)

Олег БІЛОКОНЬ
(прізвище та ініціали)

Керівник роботи _____
(підпис)

Ольга СМОТР
(прізвище та ініціали)

АНОТАЦІЯ

Олег БІЛОКОНЬ «Інформаційна система інтелектуального аналізу даних дорожньо-транспортних пригод з використанням середовища Jupyter». Дипломна робота за спеціальністю 122 «Комп'ютерні науки » складається з текстової частини, що містить 3 розділи, 73 с., 38 рис., 26 джерела.

Об'єктом дослідження дипломного проекту є сучасні інформаційні системи аналізу даних дорожньо-транспортних пригод .

Предметом дослідження дипломного проекту є аналіз інформаційних наборів даних, щодо ДТП із застосуванням інструментарію Data Science та Machine Learning для прогнозування та запобігання їх виникнення у майбутньому.

Мета роботи – дослідження методів інтелектуального аналізу даних великих обсягів з використанням інструментарію Data Science та Machine Learning, розроблення інформаційної системи інтелектуального аналізу даних дорожньо-транспортних пригод в середовищі Jupyter для прогнозування та запобігання можливості їх настання у майбутньому.

Методи дослідження – інтелектуальний аналіз Big Date, методи статистичного дослідження, алгоритми прогнозування часових рядів.

Навчальна значущість дипломного проекту полягає у набутті навичок здійснення інтелектуального аналізу даних методами Data Science та Machine Learning.

Практична значущість дипломного проекту полягає у розробленні інформаційна система інтелектуального аналізу даних дорожньо-транспортних пригод з використанням середовища Jupyter для прогнозування та запобігання можливості виникнення ДТП у майбутньому, а також для формування банку даних щодо забезпечення безпеки на дорогах.

ІНФОРМАЦІЙНО АНАЛІТИЧНА СИСТЕМА, DATA SCIENCE,
JUPYTER NOTEBOOK, DATA MINING, ДОРОЖНЬО-ТРАНСПОРТНА
ПРИГОДА

ABSTRACT

Oleh BILOKONY "Information system for the intelligent analysis of traffic accident data using the Jupyter environment". Thesis in the specialty 122 "Computer Science" consists of a text that contains 3 chapters, 73 pages, 38 figures, 26 sources.

The object of study of the diploma project is modern information systems of data analysis of traffic accidents.

The subject of the diploma project is the analysis of data sets related to traffic accidents using the tools of Data Science and Machine Learning to predict and prevent their occurrence in the future.

The purpose of the work is to research the methods of intellectual analysis of large volumes of data using the tools of Data Science and Machine Learning, to develop an information system for the intelligent analysis of traffic accident data in the Jupyter environment to predict and prevent the possibility of their occurrence in the future.

Research methods– Big Data intellectual analysis, statistical research methods, time series forecasting algorithms.

The educational significance of the diploma project lies in the acquisition of skills in intellectual data analysis using Data Science and Machine Learning methods.

The practical significance of the diploma project lies in the development of an information system for intelligent analysis of traffic accident data using the Jupyter environment to predict and prevent the possibility of road accidents in the future, as well as to form a data bank on ensuring road safety.

INFORMATION ANALYTIC SYSTEM, DATA SCIENCE, JUPYTER NOTEBOOK, DATA MINING, ROAD AND TRANSPORT ADVENTURE

ЗМІСТ

ВСТУП	Помилка! Закладку не визначено.
Розділ 1. ПРЕДМЕТНА ОБЛАСТЬ. ЗБІР ІНФОРМАЦІЇ ТА АНАЛІЗ	Помилка! Закладку не визначено.
1.1. Сучасні системи моніторингу трафіку транспортного потоку Помилка! Закладку не визначено.	
1.1.1. Системи управління мережею (Network Management Systems) Помилка! Закладку не визначено.	
1.1.2. MetroCount Traffic Executive	Помилка! Закладку не визначено.
1.1.3. Система моніторингу трафіку транспортних засобів NEXCO Central Помилка! Закладку не визначено.	
1.1.4. Система моніторингу трафіку дорожнього руху ECall Японія Помилка! Закладку не визначено.	
1.1.5. Система глобального моніторингу транспорту ECall Європа Помилка! Закладку не визначено.	
1.1.6. Система моніторингу транспортних засобів OnStar.	Помилка! Закладку не визначено.
1.1.7. Система аналізу дорожнього руху (RTA)	Помилка! Закладку не визначено.
1.1.8. HERE Traffic Analytics	Помилка! Закладку не визначено.
1.2. Порівняльний аналіз розглянутих систем	Помилка! Закладку не визначено.
Розділ 2. ДОСЛІДЖЕННЯ МЕТОДІВ ТА ЗАСОБІВ РЕАЛІЗАЦІЇ ІНФОРМАЦІЙНОЇ СИСТЕМИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДОРОЖНЬО-ТРАНСПОРТНИХ ПРИГОД	Помилка! Закладку не визначено.
2.1. Дослідження середовища розробки	Помилка! Закладку не визначено.
2.1.1. Середовище розробки Jupyter	Помилка! Закладку не визначено.
2.1.2. Бібліотека Pandas	Помилка! Закладку не визначено.
2.1.3. Бібліотека NumPy	Помилка! Закладку не визначено.
2.1.4. Бібліотека Seaborn	Помилка! Закладку не визначено.
2.1.5. Бібліотека Scikit-Learn	Помилка! Закладку не визначено.
2.2. Аналіз методів машинного навчання для розв'язання задач	

- прогнозування **Помилка! Закладку не визначено.**
- 2.2.1. Модель прогнозування на основі лінійної регресії **Помилка! Закладку не визначено.**
- 2.2.2. Модель прогнозування на основі поліноміальної регресії ... **Помилка! Закладку не визначено.**
- 2.2.3. Модель прогнозування ARIMA... **Помилка! Закладку не визначено.**
- 2.2.4. Метод експоненціального згладжування Хольта **Помилка! Закладку не визначено.**
- 2.2.5. Дерево прийняття рішень для регресії (CART) **Помилка! Закладку не визначено.**

Розділ 3. РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДОРОЖНЬО-ТРАНСПОРТНИХ ПРИГОД З ВИКОРИСТАННЯМ СЕРЕДОВИЩА JUPYTER..... **Помилка! Закладку не визначено.**

- 3.1. Вибір відповідного набору даних... **Помилка! Закладку не визначено.**
- 3.2. Дослідження факторів, що впливають на тяжкість ДТП..... **Помилка! Закладку не визначено.**
- 3.3. Прогнозування настання можливих ДТП..... **Помилка! Закладку не визначено.**
- 3.4. Створення та дослідження датасету ДТП України за 2018-2022 роки **Помилка! Закладку не визначено.**

ЗАГАЛЬНІ ВИСНОВКИ..... 8

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ..... 9

ДОДАТКИ **Помилка! Закладку не визначено.**

ЗАГАЛЬНІ ВИСНОВКИ

У магістерській дипломній роботі проведено детальний огляд та аналіз спеціалізованих систем автоматизованого моніторингу та інтелектуального аналізу трафіку дорожнього руху. На підставі проведеного аналізу встановлено, що усі вони в значній мірі сприяють підвищенню рівня безпеки на дорогах. Однак, систем автоматизованого моніторингу та інтелектуального аналізу трафіку дорожнього руху є доволі дороговартісними та надають свої послуги лише за наявності ліцензованої угоди.

Досліджено середовище розробки, алгоритмічне та математичне забезпечення, необхідне для розв'язання задачі проведення аналізу набору математичних моделей для здійснення контролю трафіку наземного автотранспорту та прогнозування виникнення можливих ДТП в часовому розрізі та їх кількості з використанням методів прогнозування Machine Learning

Як наслідок, було отримано набір інтуїтивно зрозумілих даних про кількість інцидентів настання ДТП за кожен рік або місяць, з'ясовано найбільш вагомі причини їх настання та візуалізовано їх у графічній формі тощо, а також з'ясовано прогнозовані дати настання можливих дорожньо-транспортних пригод з можливою їх кількістю на кілька місяців вперед.

На основі статистичних даних патрульної поліції України [22], оприлюднених у вільному доступі створено власний датасет, завантажено його на платформі Kaggle з метою формування банку даних щодо забезпечення безпеки на дорогах

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Слово і діло. Аналітичний портал України. «ДТП в Україні: скільки людей травмується і гине на дорогах» . [Електронний ресурс] – Режим доступу <https://www.slovoidilo.ua/2021/07/21/infografika/suspilstvo/dtp-ukrayini-skilky-lyudej-travmuetsya-hyne-dorohax> . Дата звернення: 01.12.2022
2. Українські сенсації. «Кількість смертельних ДТП на ділянках з камерами автофіксації упала втричі» . [Електронний ресурс] – Режим доступу <https://www.epravda.com.ua/news/2021/11/16/679833/>. Дата звернення: 01.12.2022
3. Центр демократії та верховенства права. «БЕЗПЕКА ДОРОЖНЬОГО РУХУ В УКРАЇНІ: ПЕРЕМОГИ ТА НОВІ ВИКЛИКИ В УМОВАХ ВІЙНИ». [Електронний ресурс]– Режим доступу <https://cedem.org.ua/news/bezpeka-dorozhnogo-ruhu-v-ukrayini/> . Дата звернення: 01.12.2022
4. Офіційний сайт ВООЗ. [Електронний ресурс]– Режим доступу <https://www.who.int/ru/news-room/fact-sheets/detail/road-traffic-injuries>. Дата звернення: 02.12.2022
5. Українські системні інновації.. «Статистика ДТП в Україні: куди рухаємося – до Європи чи до Ліберії?» . [Електронний ресурс]– Режим доступу <https://ukrsi.com.ua/news/item/?id=97> . Дата звернення: 02.12.2022
6. Офіційний сайт «Офісу трансформації». [Електронний ресурс]– Режим доступу <https://transformation.org.ua/traffic/> . Дата звернення: 12.12.2022
7. Офіційний сайт HERE Traffic . [Електронний ресурс] – Режим доступу <https://www.mbi-geodata.com/en/road-maps-here/traffic-flow/> Дата звернення: 02.12.2022
8. Офіційна документація Jupyter Notebook [Електронний ресурс] – Режим доступу <https://jupyter.org/>. Дата звернення: 15.12.2022
9. Acharya M. S., Armaan A., Antony A. S. A Comparison of Regression Models for Prediction of Graduate Admissions, 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862140.

10. Roziqin M. C., Basuki A., Harsono T. A comparison of Montecarlo linear and dynamic polynomial regression in predicting dengue fever case, 2016 International Conference on Knowledge Creation and Intelligent Computing (KCIC), Manado, 2016, pp. 213-218, doi: 10.1109/KCIC.2016.7883649.
11. Документація Pandas [Електронний ресурс] – Режим доступу <https://pandas.pydata.org/docs/index.html> . Дата звернення: 15.12.2022
12. Learning Pandas Python Tools for Data Munging, Data Analysis, and Visualization Matt Harrison. [Електронний ресурс] – Режим доступу <https://pandas.pydata.org/docs/index.html> . Дата звернення: 15.12.2022
13. Документація Seaborn. [Електронний ресурс] – Режим доступу <https://seaborn.pydata.org/examples/index.html>. Дата звернення: 20.10.2022
14. Документація Scikit-Learn. [Електронний ресурс] – Режим доступу <https://scikit-learn.org/stable/>. Дата звернення: 20.10.2022
15. Платформа Kaggle. [Електронний ресурс] – Режим доступу <https://www.kaggle.com/> . Дата звернення: 02.12.2022
16. Abu Amra I. A. S., Maghari A. Y. A. Forecasting Groundwater Production and Rain Amounts Using ARIMA-Hybrid ARIMA: Case Study of Deir El-Balah City in GAZA, 2018 International Conference on Promising Electronic Technologies (ICPET), Deir El-Balah, 2018, pp. 135-140, doi: 10.1109/ICPET.2018.00031.
17. Navarro M. M., Navarro B. B. Optimal Short-Term Forecasting Using GA-Based Holt-Winters Method, 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Macao, China, 2019, pp. 681-685, doi: 10.1109/IEEM44572.2019.8978638.
18. Python Machine Learning - Second Edition / S. Raschka, V. Mirjalili., 2017. – 622 с. – (Packt).
19. Smotr, O., Malets, R., Ljaskovska, S., Karabyn, O. (2020). Modeling of Animator Studio Control Service Functionality Using Data Mining Tools. In: Babichev, S., Peleshko, D., Vynokurova, O. (eds) Data Stream Mining & Processing. DSMP 2020. Communications in Computer and Information Science, vol 1158. Springer, Cham. https://doi.org/10.1007/978-3-030-61656-4_24

20. Miller T. Modeling Techniques in Predictive Analytics with Python and R: A Guide to Data Science / Thomas Miller. – 448 с.
21. Xia, D., et al.: A distributed WND-LSTM model on MapReduce for short-term traffic flow prediction. *Neural Comput. Appl.* 33(7), 2393–2410 (2020). <https://doi.org/10.1007/s00521-020-05076-2>
22. How Connected Vehicles Work | US Department of Transportation,” U.S. Department of Transportation, (2020). <https://www.transportation.gov/research-and-technology/how-connected-vehicles-work>, Accessed 12 May 2021
23. Kumar A. Python: Advanced Predictive Analytics / A. Kumar, J. Babcock., 2017. – 660 с. – (Packt).
24. Головатий Р.Р. Модель проекту інтелектуальної аналітичної системи обліку наукових досягнень / Р.Р. Головатий, О.О. Смотр, І.О. Малець, Н.Є. Бурак // *Středoevropský věstník pro vědu a výzkum: Central European Journal for Science and Research – Praha.* – 2019. – Вип. 3(55). – С. 88-94.
25. Сайт патрульної поліції України. Статистика ДТП. [Електронний ресурс] – Режим доступу <https://patrolpolice.gov.ua/statystyka/> Дата звернення: 05.11.2022

