УДК 004.89
О. В. Кареліна, Б. М. Липа, Р. Б. Марко, О. В. Покидко
(Тернопільський національний технічний університет ім. І. Пулюя)

ЗАСТОСУВАННЯ МОДЕЛЕЙ ГЛИБОКОГО НАВЧАННЯ ДЛЯ ВИРІШЕННЯ ЗАДАЧ КІБЕРБЕЗПЕКИ

UDC 004.89
O. V. Karelina, B. M. Lypa, R. B. Marko, O. V. Pokydko

APPLICATION OF DEEP LEARNING MODELS TO SOLVE CYBERSECURITY PROBLEMS

Моделі глибокого навчання – це моделі штучного інтелекту на базі нейронних мереж. Штучна нейронна мережа відтворює на комп'ютері роботу людського мозку за допомогою шарів нейронів. Машинне навчання – це навчання комп'ютерної нейромережі на великій кількості даних, а не згідно визначених правил. Технологія набула популярності, оскільки сучасні комп'ютери мають достатньо обчислювальних потужностей для опрацювання великих даних.

Якщо власних обчислювальних ресурсів забракне для машинного навчання, можна скористатись сервісом Google Colaboratory, що дає змогу писати й виконувати код Python у веб-переглядачі.

Машинне навчання застосовується для розв'язання задач кібербезпеки, пов'язаних з опрацюванням та аналізом великих обсягів даних: виявлення аномальних подій, підозрілої поведінки, цільових об'єктів у великому масиві даних тощо.

Нами розв'язане завдання пошуку цільової інформації (номерів кредитних карток, логінів, паролів, адрес електронної пошти тощо) у текстових масивах – парсингу сайтів, витоках корпоративних файлів та ін. Для програмної реалізації обрано мову Python через наявність спеціалізованих бібліотек опрацювання природної мови Tensorflow, Keras, NLTK.

Значним досягненням у галузі штучного інтелекту є розробка потужних моделей, які пройшли навчання на величезних обсягах даних. Для задач опрацювання мови такою моделлю є BERT (Bidirectional Encoder Representations from Transformers) [1]. BERT опубліковано у 2018 р. фахівцями машинного навчання із Google і використовується для розуміння пошукових запитів. Навчалась модель BERT на Вікіпедії та бібліотеці книг, що містить 800 мільйонів слів.

Щоб використати модель BERT для розв'язання дещо інших задач опрацювання мови, потрібно редагувати кілька нейронних шарів. Дослідники штучного інтелекту за два роки з появи моделі BERT адаптували її для різних задач. Ми скористались бібліотекою DeepPavlov [2], моделлю для класифікації текстів. В результаті із заданого масиву тексту наша модель виокремлює цільову інформацію.

Література.

1. Rani Horev. BERT Explained: State of the art language model for NLP. URL: https:// towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270 Last accessed: 30.11.2020.
2. DeepPavlov. An open source conversational AI framework. URL: https://deeppavlov.ai/ Last accessed: 30.11.2020.