Державна служба України з надзвичайних ситуацій
Львівський державний університет безпеки життєдіяльності
Навчально-науковий інститут цивільного захисту
Кафедра управління проектами, інформаційних технологій та телекомунікацій

«Допущено до захисту»
Завідувач кафедри ІТтаСЕК підполковник
служби цивільного захисту

_____ Олександр ПРИДАТКО

“____”_____20_____року

КВАЛІФІКАЦІЙНА РОБОТА
МАГІСТРА

на тему «**Розроблення системи інтелектуального аналізу даних для виявлення шахрайства в інтернет-закупівлях засобами Python**»

Виконав:
здобувач VI курсу, групи КН-61м
спеціальності (освітньої програми)
122 «Комп’ютерні науки» (Комп’ютерні науки)
(шифр і назва спеціальності (освітньої програми)
Ігор РИГАЛЬ
(ім’я та прізвище)
Керівник Ольга СМОТР
(ім’я та прізвище)
Рецензент _____
(ім’я та прізвище)

Львів – 2024 року

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 кваліфікаційну роботу

Здобувачу Ігорю РИГАЛЮ
(ім’я, прізвище)

1. Тема «Розроблення системи інтелектуального аналізу даних для виявлення шахрайства в інтернет-закупівлях засобами Python»
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 _____ 2024 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.
3. Методичні вказівки до виконання дипломної роботи магістра для здобувачів другого рівня вищої освіти спеціальності 122 «Комп’ютерні науки». Укл. Ольга Смотр, Олександр Придатко, Назарій Бурак. – Львів: Вид-во ЛДУ БЖД, 2023. – 35 с

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

Вступ

Розділ 1. Літературний огляд та постановка задач

Розділ 2. Інструменти та технології

Розділ 3. Пошук, попередня обробка і дослідницький аналіз даних

Розділ 4. Тренування моделей передбачення

Висновки

Список використаних літературних джерел

Додатки

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 кваліфікаційної роботи/проекту	Термін виконання етапів роботи	Примітка
1.	Літературний огляд та постановка задач		
2.	Інструменти та технології		
3.	Пошук, попередня обробка і дослідницький аналіз даних		
4.	Тренування моделей передбачення		

Здобувач _____ _Ігор РИГАЛЬ_____
(підпис) (прізвище та ініціали)

Керівник роботи _____ __Ольга СМОТР_____
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

Ігор РИГАЛЬ "Розроблення системи інтелектуального аналізу даних для виявлення шахрайства в інтернет-закупівлях засобами Python". Кваліфікаційна робота за спеціальністю 122 “Комп’ютерні науки” складається з ОСНОВНОЇ частини, що містить 4 розділи, с. 59, 54 рис., 12 джерел використаної літератури

Об’єкт дослідження – система інтелектуального аналізу даних для виявлення шахрайства в інтернет-закупівлях.

Предмет дослідження – технології та інструменти, що використовуються для розробки систем інтелектуального аналізу даних для виявлення шахрайства.

Мета роботи – розробити систему інтелектуального аналізу даних для виявлення шахрайства, яка буде ефективно виконувати свою ціль визначення шахрайських транзакцій.

Теоретична значущість дипломної роботи: проведено аналіз актуальності систем виявлення шахрайських схем та засобів їх реалізації, показано актуальність створення таких систем, та визначено головні переваги та недоліки деяких із популярних підходів для їх розробки.

Практична значущість дипломної роботи полягає у проектуванні та розробці моделей штучного інтелекту та порівняння їх між собою за різними метриками. Розробка велась з використанням мови програмування Python і її бібліотек. Деякі з розроблених моделей дозволяють ефективно проводити передбачення щодо законності транзакцій. Також крім тренування моделей був здійснений пошук даних і проведений дослідницький аналіз задля розуміння значущості колонок в наборі даних. Отримані результати можуть бути застосовані для покращення безпеки та ефективності інтернет-закупівельних систем, а також для запобігання фінансовим втратам компаній та користувачів.

СИСТЕМА ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ, ВИЯВЛЕННЯ ШАХРАЙСТВА, МАШИННЕ НАВЧАННЯ

ABSTRACT

Igor RYGAL "Development of an intelligent data analysis system for fraud detection in online purchases using Python". Qualifying work in specialty 122 "Computer Science" consists of a main part containing 4 sections, 59 pages, 54 figures, and 12 sources of literature used.

The object of the research is an intelligent data analysis system for fraud detection in online purchases.

The subject of the research is the technologies and tools used in the development of intelligent data analysis systems for fraud detection.

The aim of the work is to develop an intelligent data analysis system for fraud detection that will effectively fulfill its goal of identifying fraudulent transactions.

The theoretical significance of the thesis: an analysis of the relevance of fraud detection systems and the means of their implementation has been carried out, demonstrating the relevance of creating such systems, and the main advantages and disadvantages of some popular approaches for their development have been determined.

The practical significance of the thesis lies in the design and development of artificial intelligence models and their comparison based on various metrics. The development was carried out using the Python programming language and its libraries. Some of the developed models allow for effective predictions regarding the legitimacy of transactions. Additionally, besides model training, data mining was conducted, and exploratory data analysis was performed to understand the significance of columns in the dataset. The obtained results can be applied to improve the security and efficiency of online purchasing systems, as well as to prevent financial losses for companies and users.

INTELLIGENT DATA ANALYSIS SYSTEM, FRAUD DETECTION, MACHINE LEARNING

Вступ

Актуальність та новизна теми:

У сучасному світі інтернет-закупівлі стають все більш популярними, що призводить до зростання кількості шахрайства в цій сфері. Споживачі та компанії стикаються з серйозними фінансовими втратами через несанкціоновані трансакції та обманні дії. Тому розробка ефективної системи виявлення шахрайства в інтернет-закупівлях набуває великого значення. Використання різноманітних моделей машинного навчання може допомогти вирішити цю проблему.

Мета і завдання дослідження:

Метою даної магістерської роботи є розробка системи інтелектуального аналізу даних для виявлення шахрайства в інтернет-закупівлях. Для досягнення цієї мети було поставлено наступні завдання:

- Дослідження різних моделей машинного навчання для їх подальшого застосування у виявленні шахрайства.
- Розробка та реалізація системи аналізу даних для виявлення шахрайства в інтернет-закупівлях.
- Порівняльний аналіз ефективності різних моделей машинного навчання у виявленні шахрайства.

Об’єкт та предмет дослідження:

Об'єктом дослідження є процес виявлення шахрайства в інтернет-закупівлях. Предметом дослідження є розроблена система інтелектуального аналізу даних та різні моделі машинного навчання, використовані для виявлення шахрайства.

Методи дослідження:

Для досягнення поставленої мети використовувалися наступні методи:

- Логістична регресія
- Випадковий ліс
- Наївний Байєс
- Багатошаровий перцептрон
- Градієнтний бустінг

Ці методи були використані для тренування моделей на вихідних даних та оцінки їх ефективності у виявленні шахрайства.

Наукова новизна одержаних результатів:

Отримані результати є новизною у галузі виявлення шахрайства в інтернет-закупівлях. Вперше було розроблено та реалізовано систему інтелектуального аналізу даних для цієї сфери, яка використовує різні моделі машинного навчання. Результати порівняльного аналізу показали ефективність використаних моделей у виявленні шахрайства.

Публікації та апробація результатів роботи та/або практичне значення отриманих результатів:

Отримані результати можуть бути застосовані для покращення безпеки та ефективності інтернет-закупівельних систем. Вони можуть бути використані компаніями для запобігання фінансовим втратам від шахрайства. Результати дослідження можуть бути представлені на відповідних конференціях та опубліковані у наукових журналах.

ВИСНОВКИ ТА РЕКОМЕНДАЦІЇ

Проведений аналіз даних та розробка моделей для виявлення шахрайських транзакцій відкрили нові можливості для забезпечення безпеки та надійності фінансових транзакцій. Результати дозволяють зробити кілька важливих висновків, що можуть бути використані для покращення систем безпеки та зменшення ризиків.

Запропонований підхід до аналізу даних та моделювання є ефективним і може бути використаний для розпізнавання аномальних транзакцій. Кореляційний аналіз дозволяє визначити взаємозв'язки між різними факторами та їхнім впливом на можливість шахрайства.

Додатковий дослідницький аналіз виявив, що певні ознаки, такі як час, сума, категорія та місцезнаходження транзакцій, грають ключову роль у виявленні шахрайських дій. Моделі також успішно визначають та класифікують шахрайські випадки з високою точністю.

Важливою рекомендацією є впровадження та налаштування системи виявлення шахрайств на основі розробленої моделі. Додатково, рекомендується вдосконалення моделі за допомогою додаткових факторів та постійне оновлення для адаптації до змін в сфері фінансових шахрайств.

Таким чином, дана робота визначає потенціал для подальшого вдосконалення та розвитку систем безпеки фінансових транзакцій, забезпечуючи більш високий рівень надійності та захисту.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Credit Card Transactions Fraud Detection Dataset. URL: https://www.kaggle.com/datasets/kartik2112/fraud-detection/data (дата звернення: 22.12.2023)
2. PYTHON - ВИСОКОРІВНЕВА МОВА ПРОГРАМУВАННЯ. URL: https://avada-media.ua/ua/services/python-plyusy-i-minusy-yazyka-kakiye-zadachi-reshayet-i-stoit-li-izuchat/ (дата звернення: 24.12.2023)
3. Кращі IDE для Python в 2023 році. URL: https://mate.academy/blog/python/ide-for-python-2023/ (дата звернення: 24.12.2023)
4. Повний цикл дослідження даних власноруч. URL: https://medium.com/stinopys/%D0%BF%D0%BE%D0%B2%D0%BD%D0%B8%D0%B9-%D1%86%D0%B8%D0%BA%D0%BB-%D0%B4%D0%BE%D1%81%D0%BB%D1%96%D0%B4%D0%B6%D0%B5%D0%BD%D0%BD%D1%8F-%D0%B4%D0%B0%D0%BD%D0%B8%D1%85-%D0%B2%D0%BB%D0%B0%D1%81%D0%BD%D0%BE%D1%80%D1%83%D1%87-d584b01d08e6 (дата звернення: 10.01.2024)
5. Python у науці про дані: Як використовувати Python для аналізу даних та машинного навчання. URL: https://buki.com.ua/blogs/python-u-nauci-pro-dani-iak-vikoristovuvati-python-dlia-analizu-danix-ta-masinnogo-navcannia/ (дата звернення: 12.01.2024)
6. Вступ до Machine Learning: знайомство з моделями. URL: https://dou.ua/lenta/articles/introduction-machine-learning-1/ (дата звернення: 14.01.2024)
7. The Algorithmic Approach to Fraud Detection and Prevention: Insights from Leading Financial Institutions. URL: https://medium.com/@nagarjunmallesh/the-algorithmic-approach-to-fraud-detection-and-prevention-insights-from-leading-financial-b32f2046248a (дата звернення: 21.01.2024)

8. Види шахрайства в programmatic та як із ними боротися. URL: https://detector.media/production/article/141501/2018-10-04-vydy-shakhraystva-v-programmatic-ta-yak-iz-nymy-borotysya/ (дата звернення: 22.01.2024)
9. К. Ю. Кононова. МАШИННЕ НАВЧАННЯ: МЕТОДИ ТА МОДЕЛІ, 2020. 294 с.
10. Bart Baesens, Véronique Van Vlasselaer, Wouter Verbeke Ph.D. Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection, 2015. 368 c.
11. Andreas C. Müller and Sarah Guido. Introduction to Machine Learning with Python A Guide for Data Scientists, 2016. 367c.
12. Christopher M. Bishop. Pattern Recognition and Machine Learning, 2006. 729