

DOI: [10.26565/2311-2379-2021-101-04](https://doi.org/10.26565/2311-2379-2021-101-04)  
УДК 368.2:004.43

### К. Ю. КОНОНОВА

доктор економічних наук, професор  
професор кафедри економічної кібернетики та прикладної економіки  
Харківський національний університет імені В. Н. Каразіна  
майдан Свободи, 4, м. Харків, 61022, Україна  
ORCID ID: <https://orcid.org/0000-0001-6990-5746>, e-mail: [kateryna.kononova@gmail.com](mailto:kateryna.kononova@gmail.com)

### М. О. ТАРАБАНОВ

магістр кафедри економічної кібернетики та прикладної економіки  
Харківський національний університет імені В.Н. Каразіна  
майдан Свободи, 4, м. Харків, 61022, Україна  
ORCID ID: <https://orcid.org/0000-0002-2883-2162>, e-mail: [mishatarabanov98@gmail.com](mailto:mishatarabanov98@gmail.com)

## СТРАХОВІ ВИПАДКИ: АНАЛІЗ З ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ

Однією з головних проблем страхування є шахрайство, коли клієнт шляхом спотворення інформації про страховий випадок, хоче отримати завищені суми виплат. Проте традиційні методи боротьби з шахрайством у страхуванні вимагають великої кількості рутинної ручної праці та є не надто ефективними. У роботі запропоновано розробку прототипу системи моніторингу страхових випадків з метою виявлення шахрайства із застосуванням методів машинного навчання. Розробку проведено на прикладі бази даних страхових випадків, що налічує 38 змінних та містить 1000 записів зі страхових претензій клієнтів. У датасеті надана інформація щодо 1) клієнтів – 10 ознак; 2) договору страхування – 7 ознак; 3) інциденту – 21 ознака. Попередню обробку даних, побудову моделей та розробку системи моніторингу проведено з використанням мови програмування Python. На основі моделей логістичної регресії, градієнтного бустингу та випадкового лісу було побудовано низку класифікаторів з різними комбінаціями змінних. Для кожної моделі було проаналізовано матрицю спряженості, показники точності, специфічності, чутливості, побудовані ROC-криві. За результатами моделювання вдалось відібрати 5 основних змінних для моніторингу, 3 з яких характеризують клієнта, 2 – зіткнення транспортних засобів. За допомогою запропонованої системи моніторингу вдалось виявити наступні закономірності: 1) серед шахраїв найчастіше зустрічаються менеджери та працівники технічної підтримки; 2) клієнти, які у вільний час грають у шахи або займаються кросфітом, більш схильні до шахрайства; 3) більшість фактів шахрайства було зафіксовано при сильному пошкодженні транспортного засобу; 4) за відсутності контакту з екстреними службами, велика сума претензії свідчить про фрод.

Ключові слова: **страхування, виявлення шахрайства, машинне навчання, моделі класифікації, візуалізація, Python.**

*JEL Classification: C38, C55, G22.*

**Постановка проблеми.** Страхова галузь складається з величезної кількості компаній, які щорічно збирають премій на суму понад 4 трильйони доларів. Величезний розмір галузі значно збільшує ризики страхового шахрайства.

Страхове шахрайство – це протиправна поведінка суб'єктів договору страхування, внаслідок чого суб'єкти договору отримують можливість незаконно і безоплатно обертати капітал на свою користь (Пластун, 2014). Страхове шахрайство існує з часів створення самої галузі і досі залишається серйозною проблемою.

Ринок українського страхування знаходиться на стадії становлення, і саме тому на ньому так поширене шахрайство. Одним з головних чинників існування страхового шахрайства в Україні є відсутність кваліфікованих фахівців, які можуть вести ефективну боротьбу з наслідками шахрайських дій (Бондаренко, 2015). Серед інших елементів, які грають на користь злочинців, можна назвати: відсутність спеціального законодавства про страхове шахрайство; м'які вирoki в порівнянні з іншими кримінальними злочинами; низький рівень боротьби з шахрайством з боку страховиків, судів та органів прокуратури (Неджеря, 2020).

Традиційні методи боротьби з шахрайством у страхуванні вимагають великої кількості рутинної ручної праці та є не надто ефективними. Тому зараз галузь страхування зазнає стрімких змін. Компанії активно інвестують в машинне навчання, починається повноцінна гонка, в якій страховики, адаптуючись до умов конкуренції, отримують набагато якісніші портфелі договорів і задоволених клієнтів, впевнених в якості сервісу. Сьогодні машинне навчання – основний інструмент для реалізації аналітичних рішень в страхуванні. Візуалізація інформації щодо страхових випадків значно полегшує виявлення закономірностей та тенденцій у порівнянні з переглядом тисяч рядків в електронній таблиці. Навіть якщо страховий аналітик може отримувати ідеї безпосередньо з вихідних даних, діаграми та графіки спрощують передачу результатів аналізу страховому агенту чи керівництву.

**Аналіз останніх досліджень.** Розглядаючи дослідження в галузі протидії шахрайству, відмітимо статтю Пуніда «Insurance claims – Fraud detection using machine learning» (Punith, 2021), у якій автор запропонував модель, що дозволяє виявляти фрод у автострахованні методами машинного навчання. Він зазначив, що основна проблема виявлення таких випадків у незбалансованих даних (частка випадків шахрайства є відносно невеликою). Обробивши набір даних та побудувавши моделі п'яти класифікаторів, автор дійшов висновку, що найкращою моделлю є логістична регресія. Оцінивши її параметри, він спробував налаштувати гіперпараметри за допомогою інструменту GridSearchCV, який автоматично підбирає їх та створює всі можливі комбінації. В результаті автор зміг покращити базову модель та отримав ефективний інструмент для передбачення випадків шахрайства в автострахованні.

Рошан у дослідженні «Fraud Detection in Insurance Claims» (Roshan, 2021) описав проєкт по виявленню фроду у страхуванні. Спочатку він зробив підготовку даних, а саме очищення, шкалювання та кодування нечислових змінних. Далі автор виявив основні кореляції у наборі даних за допомогою статистичних методів. На відміну від попереднього дослідження, в якості основної моделі він обрав модель випадкового лісу. Автор також спробував застосувати інші ансамблеві методи, а саме бустінг з послідовною композицією алгоритмів. В підсумку він отримав модель з досить високою точністю виявлення випадків шахрайства (86%).

Щодо українських науковців, які намагались дослідити тему шахрайства у страхуванні, їх дослідження мають більш теоретичний характер. Наприклад, Ермошенко порівнює режими протидії страховому шахрайству у різних країнах на законодавчому рівні (Ермошенко, 2009). Жабинець проаналізувала зарубіжний досвід боротьби зі злочинами у страхуванні та визначила можливості впровадження відповідних методів в Україні (Жабинець, 2009). Шірінян виявила індикатори присутності шахрайських схем і запропонувала систему протидії страховому шахрайству на мікроекономічному і макроекономічному рівнях (Шірінян, 2010).

**Мета роботи** полягає в дослідженні даних страхової компанії методами машинного навчання та розробці системи моніторингу для контролю важливих змінних.

Для досягнення цієї мети було вирішено наступні **завдання**:

- проведено аналіз страхового ринку України та досягнень у боротьбі з шахрайством за допомогою методів машинного навчання;
- зібрано дані щодо страхових випадків та проведено їх попередню обробку;
- здійснено відбір значимих факторів за допомогою моделей машинного навчання та розроблено прототип системи моніторингу значущих змінних;
- проведено аналіз отриманих результатів та можливостей впровадження системи.

Попередню обробку даних, побудову моделей та розробку системи моніторингу значущих з точки зору виявлення шахрайства змінних виконано з використанням мови програмування Python.

**Основні результати дослідження.** Автостраховання є найбільш уразливою для шахраїв сферою, саме тому для розробки системи моніторингу шахрайства використано базу даних страхових випадків саме у цій галузі. База містить дані про страхові випадки та факти шахрайства за період з 1990 по 2014 рр. включно. 38 факторів, що описують страхову подію можна поділити на 3 групи:

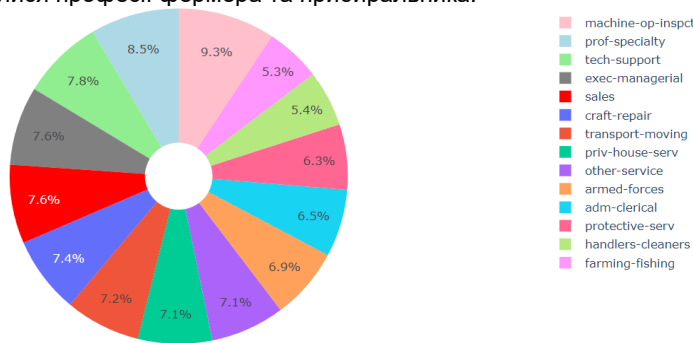
- 1) про клієнтів (вік, стать, сімейний стан, рівень освіти, хобі, термін протягом якого клієнт є споживачем страхових послуг тощо);
- 2) про договір з надавачем страхових послуг (номер полісу, ліміт страхового випадку, покриття, тощо);

3) про страхову подію (тип інциденту, тип зіткнення, кількість задіяних транспортних засобів, тілесні ушкодження, наявність свідків, залучення екстрених служб, звіт поліції, майновий позов, повідомлення про шахрайство тощо).

На першому етапі було виконано підготовку даних: виявлено та оброблено відсутні значення; виявлено та видалено викиди; категоріальні значення закодовано; проведено шкалювання даних; набір даних розділено на тестову та навчальну вибірки.

На наступному етапі на основі логістичної регресії, градієнтного бустингу та випадкового лісу було побудовано низку класифікаторів з різними комбінаціями змінних. Для кожної моделі було проаналізовано матрицю спряженості, показники точності, специфічності, чутливості, побудовані ROC-криви. Найкращою моделлю виявилась модель логістичної регресії, яка має точність 88%, чутливість 92% та специфічність 73%. За результатами моделювання вдалось відібрати п'ять основних змінних для моніторингу, три з яких характеризують клієнта, дві – зіткнення транспортних засобів, а саме: професія, захоплення й стосунки (сімейний стан) клієнта, тяжкість та тип інциденту. Розглянемо їх детальніше.

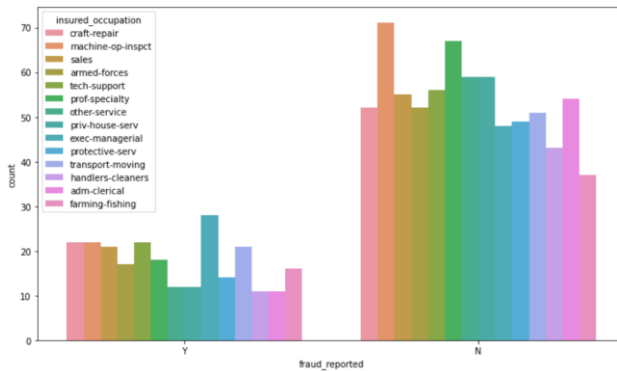
**Моніторинг професій клієнтів.** Фактор професії виявився найвпливовішим, тому розпочнемо саме з нього. Для початку розглянемо розподіл клієнтів по типу професії за допомогою гістограми, яка зображена на рис. 1. Всього у базі описано 14 професій, частотний розподіл яких є досить рівномірним, модою вибірки є професія інспектора. Найрідше серед клієнтів зустрічалися професії фермера та прибиральника.



**Рис. 1. Розподіл клієнтів по професіям**  
**Fig. 1. Distribution of clients by professions**

Джерело: авторська розробка

Звернемо увагу на розподіл професій по наявності факту шахрайства (рис. 2). Бачимо, що серед шахраїв найчастіше зустрічаються менеджери та працівники технічної підтримки. Клієнти, які працювали інспекторами, були рідше за інших помічені у скоєнні шахрайства проти страхової компанії.

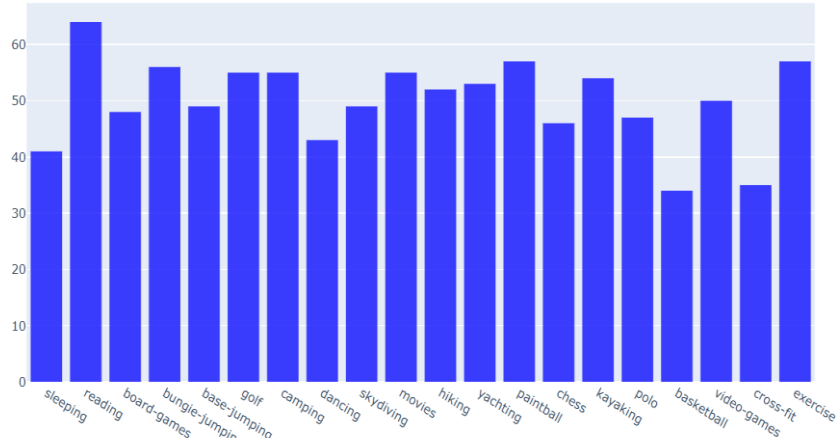


**Рис. 2. Розподіл шахраїв по професіям**  
**Fig. 2. Distribution of fraudsters by profession**

Джерело: авторська розробка

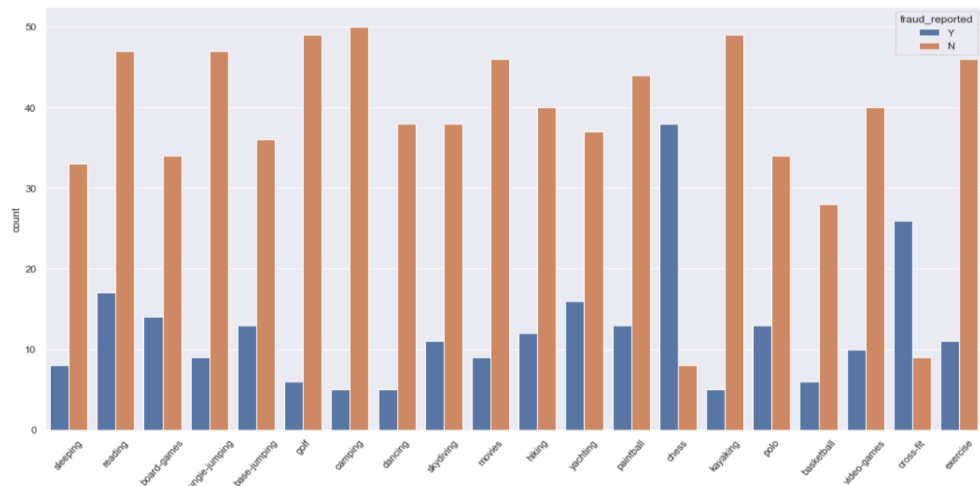
*Моніторинг захоплень клієнтів.* На рис. 3 бачимо як розподілені клієнти по варіації захоплень. Цей фактор налічує 20 унікальних значень. Найпопулярнішим захопленням серед клієнтів є читання, це значення зустрічається 63 рази. Менше за все захоплюються кросфітом та баскетболом, 34 та 36 клієнтів відповідно.

Розподіл захоплень по факту шахрайства зображено на рис. 4, тут є цікаві особливості: клієнти, які у вільний час грають у шахи або займаються кросфітом, більш схильні до шахрайства. Люди, які захоплюються туристичними походами, гольфом та каякінгом майже не були помічені у скоєнні шахрайства.



**Рис. 3. Розподіл клієнтів по захопленням**  
**Fig. 3. Distribution of customers by hobbies**

Джерело: авторська розробка

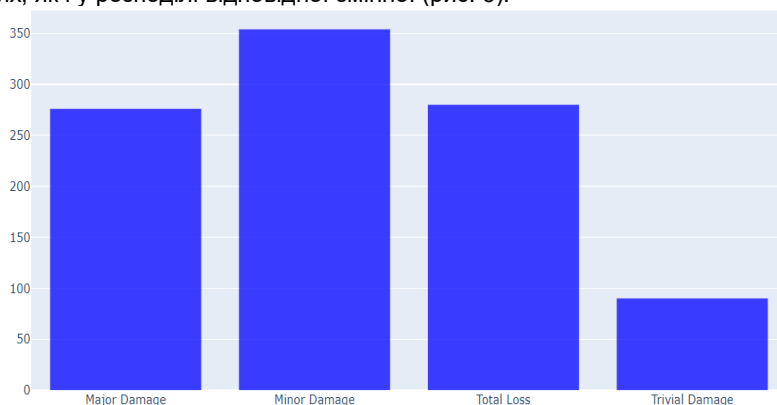


**Рис. 4. Розподіл шахраїв по захопленням**  
**Fig. 4. Distribution of fraudsters by hobbies**

Джерело: авторська розробка

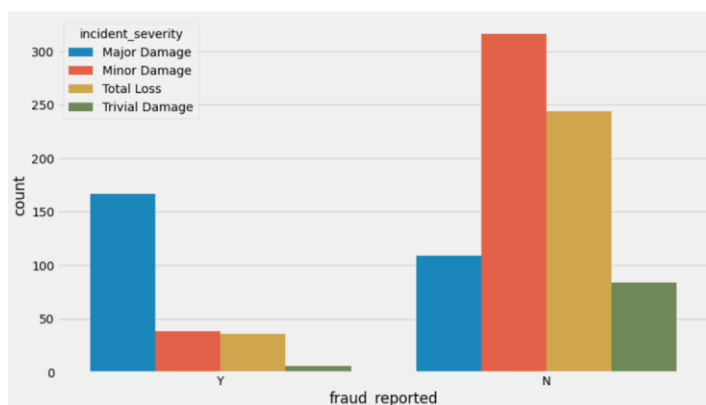
*Моніторинг шахрайства по тяжкості інциденту.* У наборі даних змінна тяжкості інциденту має 4 значення: незначне пошкодження, середнє пошкодження, сильне пошкодження та повна втрата транспортного засобу. При чому на останні три припадає переважна більшість випадків. Ситуацій незначного пошкодження транспортного засобу значно менше (рис. 5).

Найбільша кількість випадків шахрайства було зафіксовано у разі сильного пошкодження транспортного засобу. В інших інцидентах було виявлено небагато фроду та приблизно у тих же пропорціях, як і у розподілі відповідної змінної (рис. 6).



**Рис. 5. Розподіл інцидентів по тяжкості наслідків**  
**Fig. 5. Distribution of incidents by severity of consequences**

*Джерело: авторська розробка*



**Рис. 6. Розподіл випадків шахрайства по тяжкості інцидентів**  
**Fig. 6. Distribution of fraud cases by severity of incidents**

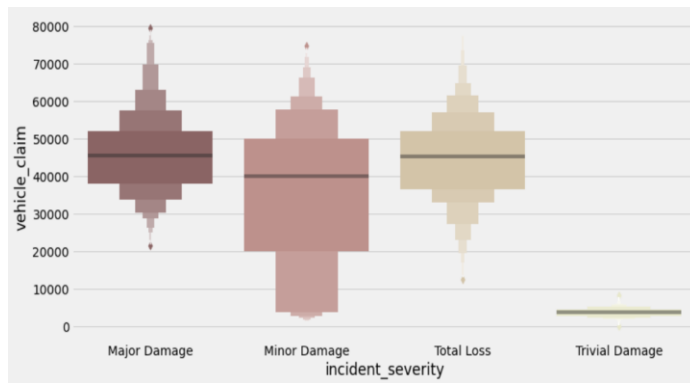
*Джерело: авторська розробка*

На рис. 7 наведено залежність між претензіями клієнтів на виплату грошей за пошкоджений транспортний засіб та тяжкістю інциденту. Можемо спостерігати, що суми претензій при середньому пошкодженні вищі ніж при повній втраті транспортного засобу, це вказує на наявність шахрайства з боку клієнтів. Щодо незначного пошкодження транспортного засобу, то тут фроду майже немає, тому що в цьому випадку важко інсценувати суттєві наслідки та претендувати на великі страхові виплати.

**Моніторинг шахрайства по типу інциденту.** Тип інциденту також має 4 значення: зіткнення одного або кількох транспортних засобів, пошкодження припаркованого засобу пересування або його викрадення. У наборі даних переважають 2 типи значень – одиночне зіткнення та зіткнення кількох транспортних засобів (рис. 8).

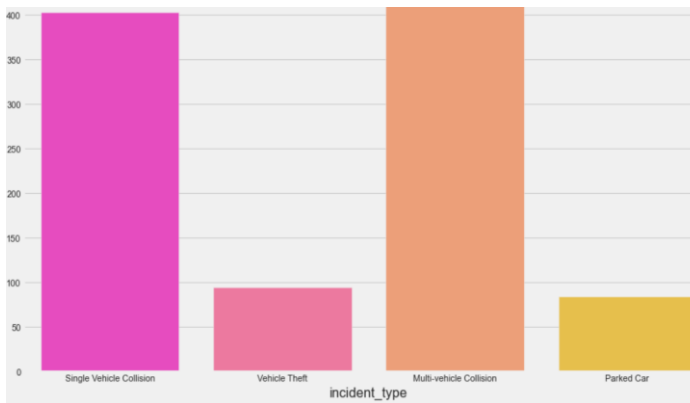
Саме у цих двох випадках ознаки фроду виявлялися найчастіше (рис. 9). В двох інших випадках, шахрайства майже не було, мабуть тому що їх більш важко інсценувати, це занадто ризиковано для недоброчесних клієнтів.

Крім означених вище, є ще кілька ознак (не значущих за p-value логістичної регресії), яким важливо приділити увагу.



**Рис. 7. Претензії на транспортний засіб**  
**Fig. 7. Claims on the vehicle**

Джерело: авторська розробка



**Рис. 8. Розподіл страхових випадків по типу зіткнення**  
**Fig. 8. Distribution of insured events by type of collision**

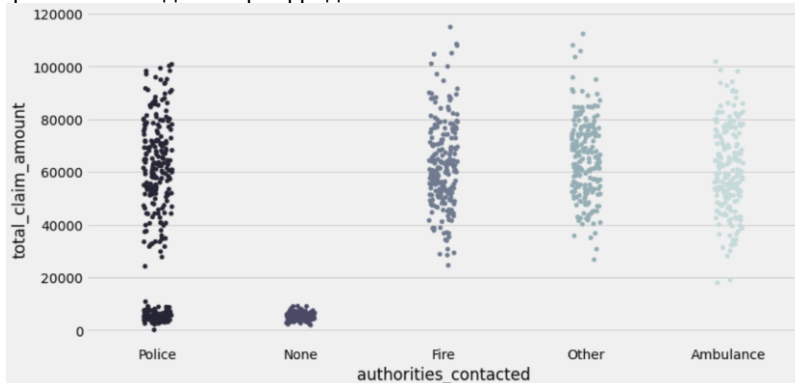
Джерело: авторська розробка



**Рис. 9. Розподіл шахрайства по типу зіткнення**  
**Fig. 9. Distribution of fraud by type of collision**

Джерело: авторська розробка

**Контакт з екстреними службами.** На рис. 10 можемо спостерігати, як співвідносяться факт контакту з екстреними службами (поліція, пожежна охорона, швидка медична допомога та інші) та сума виплати по страховій події. Бачимо, що у разі контакту з будь-якою службою у позиках було зазначено велику суму виплати від страхової компанії. Ознаки шахрайства у цих випадках визначити досить важко за допомогою машинного навчання, тут потрібна експертна думка страхового агента. Але якщо контакту з екстреними службами не було, то велика сума претензії скоріш за все свідчить про фрод.



**Рис. 10. Залежність суми виплат від контакту з екстреними службами**  
**Fig. 10. Dependence of the amount of payments on contact with emergency services**

Джерело: авторська розробка

**Відношення між сплатою щорічних внесків та сумою виплат по страховим подіям.** Для аналізу цього фактору було побудовано діаграму розсіювання (рис. 11). Червоними точками виділені страхові випадки добросовісних клієнтів, синіми з елементами шахрайства. Бачимо, що шахрайські випадки зосереджені переважно у правому кластері (приблизно від 40 000\$ до 90 000\$). Тобто, частіше у фроді були завищені суми претензій. Значення, які знаходяться у лівій частині діаграми описують вигідних клієнтів, їх страхові виплати були невеликими (в цьому діапазоні знаходиться приблизно 50% клієнтів). Друга половина клієнтів є збитковими для страхової компанії. Проте в цілому у цей кластер потрапили переважно добросовісні клієнти.



**Рис. 11. Залежність щорічних внесків клієнтів та сум виплат**  
**Fig. 11. Dependence of annual customer contributions and payment amounts**

Джерело: авторська розробка

**Висновки.** Технології машинного навчання набирають все більшої популярності серед страхових компаній та трансформують всі аспекти цієї галузі. Існує потенціал для високої окупності інвестицій в автоматизовані системи виявлення шахрайства, оскільки такі технології можуть безпосередньо впливати на усунення шахрайства у сфері розгляду претензій та зменшити збитки страхової компанії. Тож зрештою страхові компанії стикаються з вибором: або покрити витрати на впровадження нових технологій виявлення шахрайства сьогодні, або втрачати прибуток через шахрайство, сподіваючись, що найближчим часом послуги виявлення шахрайства подешевшають. Згідно зі статистикою Федерального Бюро Розслідувань США, 80% організацій зазначили, що головною проблемою впровадження нових технологій є саме фінансування, тому насамперед страховій компанії важливо оцінити результати від впровадження системи моніторингу шахрайства (FBI, 2020).

Для розробки системи моніторингу шахрайства в роботі було використано базу, що містить дані про страхові випадки та факти шахрайства у автострахованні. За результатами моделювання вдалось відібрати 5 основних змінних для моніторингу, 3 з яких характеризують клієнта, 2 – зіткнення транспортних засобів. З використанням методів машинного навчання вдалось виявити наступні цікаві закономірності: 1) серед шахраїв найчастіше зустрічаються менеджери та працівники технічної підтримки, 2) клієнти, які у вільний час грають у шахи або займаються кросфітом, більш схильні до шахрайства, 3) більше всього фактів шахрайства було зафіксовано при сильному пошкодженні транспортного засобу, у випадках одиночного зіткнення та зіткнення кількох транспортних засобів, 4) за відсутності контакту з екстремними службами, велика сума претензії свідчить про фрод.

Розроблений прототип моніторингу може бути використаний в якості системи підтримки прийняття рішень в сфері шахрайства у страхуванні. Завдяки запропонованим варіантам візуалізації значущих факторів страхові компанії зможуть швидше розпізнавати факти шахрайства. Вдосконалення запропонованої системи моніторингу можливо насамперед за рахунок аналізу більш широкого спектра даних. Цікавим напрямком розвитку системи є можливість використання штучного інтелекту для роботи з неструктурованими даними. Наприклад, аналіз зображень допоможе визначити, чи були підроблені фотографії аварій, подані заявниками. Технології інтелектуального аналізу текстів допоможуть аналізувати претензії, виявляти тенденції та закономірності, пов'язані з певними ключовими словами.

#### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Пластун В.Л. Проблеми страхового шахрайства та практика його уникнення. *Економіка: проблеми теорії та практики*. Дніпропетровськ : ДНУ, 2014. С. 477–488.
2. Бондаренко Є.В. Злочини в сфері страхування: особливості вчинення їх в Україні. *Електронний репозитарій НАВС*. 2015. URL: <http://elar.naiu.kiev.ua/jspui/handle/123456789/8505>.
3. Неджеря В.І. Ризики страхового шахрайства та методи боротьби з ними. *Ефективна економіка*. 2020. № 3. DOI: <https://doi.org/10.32702/2307-2105-2020.3.150>.
4. Punith A. Insurance claims – Fraud detection using machine learning. 2021. URL: <https://medium.com/geekculture/insurance-claims-fraud-detection-using-machine-learning-78f04913097>.
5. Roshan S. Fraud Detection in Insurance Claims. 2021. URL: <https://www.kaggle.com/roshansharma/fraud-detection-in-insurance-claims/notebook#Modelling-with-Ensemble-of-Samplers>.
6. Єрмошенко А.М. Страхове шахрайство як джерело виникнення загроз у взаємодії страховиків і банків. *Збірник наукових праць ДВНЗ „Українська академія банківської справи НБУ”*. 2009. Вип. 27. URL: <http://essuir.sumdu.edu.ua/handle/123456789/54055e>.
7. Жабинець О.Й. Попередження страхових зловживань як один з факторів забезпечення економічної безпеки страховика. *Науковий вісник Львівського державного університету внутрішніх справ*. 2009. №1. С. 1–6. URL: [https://www.lvduvs.edu.ua/documents\\_pdf/visnyky/nvse/01\\_2009/09zojeb.pdf](https://www.lvduvs.edu.ua/documents_pdf/visnyky/nvse/01_2009/09zojeb.pdf)
8. Шірінян Л.В. Страхове шахрайство – економіко-правові аспекти, індикатори і шляхи боротьби. *Економіка та право*. 2010. № 3. URL: <http://dspace.nuft.edu.ua/jspui/handle/123456789/16581>.
9. Insurance Fraud. *Federal Bureau of Investigation (FBI)* : веб-сайт. URL: <https://www.fbi.gov/stats-services/publications/insurance-fraud>.

Стаття надійшла до редакції 03.10.2021

Стаття рекомендована до друку 22.11.2021



## REFERENCES

1. Plastun, V. (2014). Problems of insurance fraud and the practice of avoiding it. *Economics: problems of theory and practice*, 477-488. (in Ukrainian)
2. Bondarenko, E. (2020). Crimes in the field of insurance: features of their commission in Ukraine. *Electronic repository of NAVS*, 3. Retrieved from <http://elar.naiu.kiev.ua/jspui/handle/123456789/8505>. (in Ukrainian)
3. Nedzherya, V. (2020). Risks of insurance fraud and methods of combating them. *Efficient economy*, 3. doi: <https://doi.org/10.32702/2307-2105-2020.3.150>. (in Ukrainian)
4. Punith, A. (2021). Insurance claims – Fraud detection using machine learning. Retrieved from <https://medium.com/geekculture/insurance-claims-fraud-detection-using-machine-learning-78f04913097>.
5. Roshan, S. (2021). Fraud Detection in Insurance Claims. Retrieved from <https://www.kaggle.com/roshansharma/fraud-detection-in-insurance-claims/notebook#Modelling-with-Ensemble-of-Samplers>.
6. Ermoshenko, A. (2009). Insurance fraud as a source of threats in the interaction of insurers and banks. *Ukrainian Academy of Banking of the NBU*, 27. Retrieved from <http://essuir.sumdu.edu.ua/handle/123456789/54055e>. (in Ukrainian)
7. Zhabynets, O. (2009). Prevention of insurance abuse as one of the factors ensuring economic security of the insurer. *Bulletin of Lviv State University of Internal Affairs*, 1, 1-6 Retrieved from [https://www.lvduvs.edu.ua/documents\\_pdf/visnyky/nvse/01\\_2009/09zojebs.pdf](https://www.lvduvs.edu.ua/documents_pdf/visnyky/nvse/01_2009/09zojebs.pdf). (in Ukrainian)
8. Shirinyan, L. (2010). Insurance fraud – economic and legal aspects, indicators and ways to fight. *Economics and law*, 3. Retrieved from <http://dspace.nuft.edu.ua/jspui/handle/123456789/16581>. (in Ukrainian)
9. FEDERAL BUREAU OF INVESTIGATION (FBI). 2020. Insurance Fraud. Retrieved from <https://www.fbi.gov/stats-services/publications/insurance-fraud>.

The article was received by the editors 03.10.2021

The article is recommended for printing 22.11.2021

**K. KONONOVA**, D.Sc. (Economics), Professor, Professor of the Department of Economic Cybernetics and Applied Economics, V.N. Karazin Kharkiv National University  
4 Svobody Sq., Kharkiv, 61022, Ukraine  
ORCID ID: <https://orcid.org/0000-0001-6990-5746>, e-mail: [kateryna.kononova@gmail.com](mailto:kateryna.kononova@gmail.com)  
**M. TARABANOV**, Master of the Department of Economic Cybernetics and Applied Economics,  
V.N. Karazin Kharkiv National University  
4 Svobody Sq., Kharkiv, 61022, Ukraine  
ORCID ID: <https://orcid.org/0000-0002-2883-2162>, e-mail: [mishatarabanov98@gmail.com](mailto:mishatarabanov98@gmail.com)

## INSURANCE CASES: ANALYSIS BY MACHINE LEARNING

One of the main problems of insurance is fraud, when the client wants to get overpayments by distorting information about the insured event. However, traditional methods of insurance fraud combating require a lot of routine manual work and are not very effective. The paper proposes the development of a prototype of the insurance case monitoring system in order to detect fraud using machine-learning methods. The development was carried out on the example of a database of insurance cases, which has 38 variables and contains 1000 records of insurance claims. The dataset provides information on 1) client – 10 features; 2) insurance contract – 7 features; 3) incident – 21 features. Preliminary data processing, modeling and development of the monitoring system was carried out using the Python. Classifiers (logistic regression, gradient boosting and random forest) with different combinations of variables were built. For each model, the conjugation matrix, accuracy, specificity, sensitivity, and ROC curves were analyzed. Simulation results allowed to select 5 main variables for monitoring, 3 of which characterize the client, 2 – incident. The proposed monitoring system allows to identify the following patterns: 1) in most cases, fraudsters were managers and technical support staff; 2) customers, who were practicing chess or CrossFit, were more prone to fraud; 3) most of the fraud was recorded in severe damage; 4) in case of absence of contact with emergency services, a large amount of the claim indicated fraud.

Keywords: **insurance, fraud detection, machine learning, classifiers, visualization, Python.**

JEL Classification: C38, C55, G22.

**Е. Ю. КОНОНОВА**, доктор економічних наук, професор, професор кафедри економічної кібернетики і прикладної економіки, Харківський національний університет імені В. Н. Каразіна  
пл. Свободи, 4, г. Харків, 61022, Україна

ORCID ID: <https://orcid.org/0000-0001-6990-5746>, e-mail: [kateryna.kononova@gmail.com](mailto:kateryna.kononova@gmail.com)

**М. А. ТАРАБАНОВ**, магістр кафедри економічної кібернетики і прикладної економіки,  
Харківський національний університет імені В. Н. Каразіна  
пл. Свободи, 4, г. Харків, 61022, Україна

ORCID ID: <https://orcid.org/0000-0002-2883-2162>, e-mail: [mishatarabanov98@gmail.com](mailto:mishatarabanov98@gmail.com)

## СТРАХОВЫЕ СЛУЧАИ: АНАЛИЗ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

Одной из главных проблем страхования является мошенничество, когда клиент путем искажения информации о страховом случае хочет получить завышенные суммы выплат. Однако традиционные методы борьбы с мошенничеством в страховании требуют больших объемов рутинной ручной работы и не слишком эффективны. В работе предложена разработка прототипа системы мониторинга страховых случаев с целью выявления мошенничества с применением методов машинного обучения. Разработка проведена на примере базы данных страховых случаев, которая насчитывает 38 переменных и содержит 1000 записей по страховым претензиям клиентов. В датасете представлена информация про 1) клиента – 10 признаков; 2) договор страхования – 7 признаков; 3) инцидент – 21 признак. Предварительная обработка данных, построение моделей и разработка системы мониторинга проведены с использованием языка программирования Python. На основе моделей логистической регрессии, градиентного бустинга и случайного леса построен ряд классификаторов с разными комбинациями переменных. Для каждой модели была проанализирована матрица сопряженности, показатели точности, специфичности, чувствительности, построены ROC-кривые. По результатам моделирования удалось отобрать 5 основных переменных для мониторинга, 3 из которых характеризуют клиента, 2 – столкновение транспортных средств. С помощью предложенной системы мониторинга удалось выявить следующие закономерности: 1) среди мошенников чаще всего встречаются менеджеры и работники технической поддержки; 2) клиенты, в свободное время играющие в шахматы или занимающиеся кроссфитом, более склонны к мошенничеству; 3) большинство фактов мошенничества было зафиксировано при сильном повреждении транспортного средства; 4) при отсутствии контакта с экстренными службами, большая сумма претензии свидетельствует о фроде.

Ключевые слова: **страхование, выявление мошенничества, машинное обучение, модели классификации, визуализация, Python.**

*JEL Classification: C38, C55, G22.*

---

**Як цитувати:** Кононова, К. Ю., & Тарабанов, М. О. (2021). Страхові випадки: аналіз з використанням машинного навчання. *Вісник Харківського національного університету імені В. Н. Каразіна серія «Економічна»*, (101), 35-44. <https://doi.org/10.26565/2311-2379-2021-101-04>.

**In cites:** Kononova, K., & Tarabanov, M. (2021). Insurance cases: analysis by machine learning. *Bulletin of V. N. Karazin Kharkiv National University Economic Series*, (101), 35-44. <https://doi.org/10.26565/2311-2379-2021-101-04>. (in Ukrainian)

---