

ІДЕНТИФІКАЦІЯ БІЗНЕС-МОДЕЛЕЙ СТРАХОВИКІВ В УКРАЇНІ: КЛАСТЕРНИЙ АНАЛІЗ ТА МАШИННЕ НАВЧАННЯ

ОЛЕКСАНДР ТАРНАВСЬКИЙ^{ab}, ВІКТОР КОЛОМІЄЦЬ^a

^aНаціональний банк України
E-mail: Oleksandr.Tarnavskiy@bank.gov.ua
Viktor.Kolomiets@bank.gov.ua

^bНаціональний університет “Кієво-Могилянська академія”

Анотація

У цьому дослідженні розглядається діяльність компаній зі страхування іншого, ніж страхування життя, що функціонували на ринку України у 2019–2020 роках. Зокрема, використано набір методів кластеризації – класичний алгоритм K-середніх і самоорганізаційні карти Кохонена, щоб дослідити характеристики роздрібної, корпоративної, універсальної (представлена двома кластерами) та перестрахової бізнес-моделей. Результати кластеризації валідовано за допомогою класичних показників та показника міграції, що забезпечує стабільність кластерів у часі. Проаналізовано міграцію компаній між виділеними кластерами (зміна бізнес-моделі) протягом досліджуваного періоду та виявлено значну міграцію між перестраховою та корпоративною моделями, а також у межах універсальної моделі. Аналіз даних про припинення поточної діяльності страховиків дає змогу зробити висновок, що компанії, які дотримуються універсальної бізнес-моделі, виявляються найбільш фінансово стабільними, а їхні аналоги, об’єднані в перестраховий кластер, імовірно, залишаються найменш стабільними. Результати цього дослідження можуть бути корисними для страхового нагляду та мати значний вплив на регуляторну політику.

Класифікація JEL G22, D22

Ключові слова кластерний аналіз, нейронні мережі, бізнес-модель, страхування

1. ВСТУП

З першого липня 2020 року НБУ почав упровадження реформи фінансового сектору, поширюючи вимоги прозорості, надійності та ефективності на небанківський фінансовий сектор. Першочерговою метою реформ є підвищення якості послуг зі страхування та захист інтересів страхувальників.

Для ефективного нагляду, контролю та запровадження реформ небанківського фінансового ринку потрібне розуміння структури ринку і того, як ведуть свій бізнес його учасники. Наприклад, різні бізнес-моделі можуть мати досить різні профілі ризику. Виділення однорідних груп компаній зі схожими ризиками дає змогу більш детально проаналізувати стійкість і платоспроможність страхових компаній та ефективно передбачати кризові події. Це дослідження має на меті сприяти розумінню структури українського страхового ринку, а також операційної діяльності і профілю ризиків компаній. У

ньому визначено бізнес-моделі українських страховиків та їхні ключові особливості з використанням кількісних показників для сприяння нагляду за страховим ринком.

Щоб досягти поставленої мети, ми намагаємося відповісти на низку запитань. Чи можна ідентифікувати однорідні та стабільні групи (бізнес-моделі) страхових компаній на основі нормативних даних? Які ключові характеристики цих бізнес-моделей? Як компанії змінили свої бізнес-моделі за період дослідження? Чи можуть деякі бізнес-моделі бути пов’язаними з підвищеними ризиками?

Для цього дослідження ми провели кластерний аналіз українського страхового ринку, щоб визначити бізнес-моделі, які використовують страховики. Ми застосовуємо низку методів кластеризації, включно з ієрархічними, неієрархічними та методами машинного навчання. На основі методу K-середніх виділяємо п’ять кластерів, що відповідають чотирьом бізнес-моделям – корпоративній, роздрібній, універсальній (ділиться на

два кластери) та перестраховій. Перед застосуванням алгоритмів кластеризації було виділено штучний кластер “неактивна” (до нього увійшли компанії, що були малоактивними або взагалі не займалися страховою діяльністю, проте мали ліцензію і увійшли до вибірки). Оцінити якість проведеного нами дослідження допомагає низка обрахованих коефіцієнтів, зокрема коефіцієнт міграції та коефіцієнт силуєту.

Ми аналізуємо бізнес-моделі як за допомогою характеристик, на підставі яких відбувалося розділення, так і використовуючи додаткові змінні, які не використовуються в алгоритмах кластеризації. Так, компанії бізнес-моделі корпоративна страхують здебільшого юридичних осіб, моделі роздрібна – навпаки, фізичних. Компанії з універсальною бізнес-моделлю широко використовують офіси як канал продажу, тоді як з перестраховою – не використовують їх взагалі. Детальніший опис кластерів наведено в тексті.

Далі ми показуємо, як відбувалася міграція страховиків між кластерами в період з 2019 до 2020 року. Ми спостерігаємо значну міграцію між бізнес-моделями перестрахова та корпоративна і всередині бізнес-моделі універсальна. Також спостерігаємо значну міграцію до штучного кластера неактивні, тобто припинення компаніями страхової діяльності. На підставі цих міграцій можна емпірично зробити висновки про рівень ризику конкретної бізнес-моделі. Так, найбільша частка компаній, що вийшли з ринку за досліджуваній період, належали до бізнес-моделі перестрахова; більше половини компаній цього кластера вийшли з ринку у 2020 році. Суттєва міграція до групи неактивні також спостерігалася для бізнес-моделей корпоративна і роздрібна.

Стаття має таку структуру. У другому розділі подано огляд актуальної літератури. У третьому розділі докладно описані методологія, дані та програмне забезпечення, які застосовувалися в цьому аналізі. У четвертому розділі наведено ключові результати дослідження, а саме опис бізнес-моделей та міграція між ними. У четвертому розділі представлені ключові результати дослідження та показано ризикованість кожної з ідентифікованих бізнес-моделей. Висновки стисло підсумовують результати дослідження та окреслюють перспективні напрями подальшого розроблення тематики.

2. ОГЛЯД ЛІТЕРАТУРИ

Розвитку досліджень бізнес-моделей банків сприяли Базельські рекомендації та впровадження процесу SREP. Дослідження бізнес-моделей страховиків є також перспективним напрямом досліджень.

Більша частина наукових праць спрямована на аналіз та сегментацію клієнтів страхових компаній. Дослідження Wang and Keogh (2008), Zaquie (2019) присвячені кластерному аналізу для визначення цільової аудиторії компанії. Методи кластеризації використовувалися для визначення портрета клієнтів на підставі наборів даних, отриманих із трансакцій полісів і профілів власників полісів. Праця Вана та Кіо реалізовувалася за допомогою карт самоорганізації (SOM) та алгоритму K-середніх. Метод K-середніх також використано в публікації Abolmakarem et al. (2016), спрямованій на сегментацію з метою визначення найбільш прибуткових клієнтів для компанії. Velykoivanenko and Beschastna (2018) використовували карти Кохонена, щоб класифікувати

страхові компанії з погляду їх фінансових показників на три групи. Потім вони об'єднали результати кластеризації з рейтингом експертів, щоб отримати інтегральний показник фінансової стабільності кожної компанії.

Більшість авторів використовують поєднання двох методів кластеризації. Зокрема, у дослідженні, проведеному Kramarić et al. (2018), європейські страхові компанії розподілені на сім кластерів. На відміну від попередніх досліджень, тут кластеризуються компанії, а не їхні клієнти. За допомогою поєднання ієрархічної кластеризації та кластеризації методом K-середніх 119 страховиків поділяються на сім груп залежно від країни походження та типу компанії. Bach et al. (2020) використовують карту Кохонена в поєднанні з ієрархічним кластерним аналізом для дослідження ризику шахрайства лізингової компанії. Нейрони карти Кохонена поєднують у п'ять кластерів за допомогою методу Уорда, після чого аналізують ризик-характеристики кластерів.

Праця Ahmar et al. (2018) є прикладом кластерного аналізу поза сферою фінансів. У своєму дослідженні автори за допомогою методу K-середніх групують провінції Індонезії. Таке групування, на їхню думку, має допомогти класифікувати райони країни для легшого подолання соціальних проблем. Abbas et al. (2020) порівняли методи K-середніх та K-медоїд, використавши дані про жінок у період вагітності. Метод K-медоїд за своєю суттю дуже схожий на метод K-середніх, тому часто застосовується в поєднанні з ним. Авторі показують, що для конкретних даних метод K-медоїд є більш точним, ніж K-середніх.

Rashkovan and Pokidin (2016) здійснили ідентифікацію бізнес-моделей банків в Україні за допомогою карти самоорганізації Кохонена та провели паралелі між бізнес-моделями та показниками різних типів ризику, на які банк може наражатися. За методологією наша праця найбільше подібна саме до цього дослідження. Проте, на відміну від Рашкована та Покідіна, які будують висновки свого дослідження на карті самоорганізації Кохонена, ми використовуємо цей метод як додатковий до методу K-середніх. Також кластерний аналіз із метою виділення бізнес-моделей використали Ferstl and Seres (2012). На відміну від попередніх дослідників, вони використали алгоритм K-середніх, що ґрунтується на використанні відстані Махаланобіса. У цій праці авторам вдалося виділити п'ять бізнес-моделей банків на основі п'яти їхніх показників.

Більшість авторів використовують найпростіші моделі кластеризації, серед яких метод K-середніх. Ми у своїй праці спробували розробити методологію, яка допомагає визначити розподіл компаній відповідно до їхньої бізнес-моделі. Для реалізації дослідження ми використали широкий набір інструментів кластеризації, проте висновки будували на методі K-середніх. Самоорганізаційні карти Кохонена в нашій роботі є зручним інструментом візуалізації. У цій праці оцінка якості кластеризації вперше ґрунтується на використанні коефіцієнта міграції – показника, що характеризує стабільність кластерів.

3. ДАНІ ТА МЕТОДОЛОГІЯ

Опис використаних даних

Для проведення кластерного аналізу страхових компаній ми збрали дані з регуляторних звітів

Таблиця 1. Показники, за якими відбувався поділ на кластери

№ з/п	Назва показника	Назва змінної	Формула
1	Рентабельність активів	ROA	Чистий прибуток / вартість активів
2	Кількість представництв у регіонах	Offices	Загальна кількість офісів, що використовуються та не є головним офісом
3	Частка премій з обов'язкових видів страхування в загальному обсязі зібраних премій	% of mandatory	Обсяг премій з обов'язкових видів страхування / загальний обсяг премій
4	Частка премій від юридичних осіб	Corporate	Обсяг премій від юридичних осіб / загальний обсяг премій
5	Частка вхідного перестраховування у преміях	Re-to-premiums	Обсяг премій з перестраховування / загальний обсяг премій

247 українських страховиків за два роки – 2019 та 2020 роки. За аналізований період кількість діючих страховиків суттєво зменшилася. Так, на кінець 2020 року база даних складалася зі спостережень щодо 185 страховиків. Дані було зібрано з бази даних звітів, що страховики подають регулятору.

Для визначення бізнес-моделі ми намагалися відібрати показники, які б відповідали на такі питання щодо компанії:

- на яких клієнтів орієнтується?
- на яких видах страхування концентрується та наскільки очевидно?
- які канали продажу використовує?

Ми обрали показники, які водночас допомагали би отримати відповіді на ці питання та за допомогою яких розподіл на кластери відбувався б оптимально з огляду на кількісні метрики – коефіцієнт міграції та коефіцієнт силуєту (описано далі в тексті). Відповідно до отриманих значень цих метрик був обраний оптимальний набір показників для кластеризації. Окремо було виділено різні набори показників, які дають змогу ширше описати кластери та ризики, що їм притаманні. Ці показники не були включені в модель, оскільки розділення на їхній основі призводило до гірших показників якості кластеризації. Навпаки, вони були використані для ширшої характеристики визначених кластерів. Таблиці 1 і 2 описують алгоритм розрахунку цих характеристик.

Після розрахунку показників їхні значення були стандартизовані (середнє значення приведені до 0, дисперсія – до 1). Це потрібно з огляду на чутливість алгоритмів кластеризації до дисперсії в даних. Ми також визначили викиди в даних. Спостереження, що містяться далі трьох стандартних відхилень від середнього, були приведені до найближчих значень, що входять у діапазон трьох стандартних відхилень. Розподіл спостережень до та після цієї процедури наведено на рисунку Б.2.

Компанії, загальний обсяг премій яких за звітний період не перевищував 5 млн грн, віднесено до штучного кластера (групи) – “неактивні” у звітному періоді. Загальна частка на ринку (у преміях) таких компаній у 2020 році становила менше 1%.

Нижче наводиться описова статистика даних за 2020 рік.

Наведені дані свідчать, що більшість компаній у 2020 році були малоприбутковими або збитковими, на відміну від вищих рівнів прибутковості, які спостерігалися в попередні роки (таблиця А.1). Причиною такого падіння прибутковості можна назвати підвищення рівня вимог до медичного страхування як наслідок COVID-19. Наприклад, резерв збитків медичного страхування значно зростає у періоди сплесків нових випадків COVID-19.

Більша частина ринку зосереджує свою діяльність на дешевих договорах, які, найімовірніше, продають

Таблиця 2. Показники, використані для додаткового опису кластерів

№ з/п	Назва показника	Назва змінної	Формула
1	Відношення частки перестраховиків у страхових резервах до загального обсягу страхових резервів	Re-to-provisions	Сума прав вимоги до перестраховиків / сума страхових резервів
2	Показник збитковості	Loss ratio	(Страхові Виплати + витрати, пов'язані з регулюванням страхових випадків + зміна в резервах збитків) / (премії + зміна резерву незароблених премій)
3	Середній розмір зібраної премії	Mean premium	Сума зібраних премій / Кількість договорів страхування
4	Відношення заробітної плати до зібраних премій	Wages/Premiums	Сума витрат на заробітну плату / сума зібраних премій за звітний період
5	Максимальний рівень концентрації у групі видів страхування	Concentration	Максимальне значення обсягу премій серед 7 категорій* / загальний обсяг премій

* Перелік категорій: 1) ядерне страхування; 2) транспортне інше; 3) автомобільне; 4) відповідальності; 5) особисте; 6) майнове; 7) інше.

Таблиця 3. Описова статистика змінних моделі

	ROA	Offices	Re-to-premiums	% of mandatory premiums	Corporate
Середнє	0.02	8.16	0.08	0.19	0.51
Ст. відхилення	0.10	24.99	0.21	0.26	0.33
Мін.	-0.67	0.00	0.00	0.00	0.00
Q (25%)	0.00	0.00	0.00	0.00	0.22
Q (50%)	0.01	0.00	0.00	0.03	0.46
Q (75%)	0.05	0.00	0.03	0.35	0.86
Макс.	0.40	200.00	1.00	0.86	1.00

Таблиця 4. Описова статистика інформативних показників компанії

	Re-to-provisions	Loss ratio	Mean premium	Wages/Premiums	Concentration
Середнє	0.21	0.39	116.41	0.06	0.68
Ст. відхилення	0.24	0.51	405.89	0.06	0.19
Мін.	0.00	-0.80	0.00	0.00	0.29
Q (25%)	0.02	0.11	0.89	0.02	0.54
Q (50%)	0.12	0.35	3.22	0.04	0.66
Q (75%)	0.36	0.52	29.79	0.07	0.83
Макс.	0.93	4.08	3,534.63	0.46	1.00

фізичним особам. Медіанний “середній чек” – близько 3 тис. грн.

За нашими даними, більшість компаній не мала офісів. З одного боку, це може свідчити про переважання на ринку бізнес-моделей, які не використовують офіси як канал продажу. З іншого боку, таке сильне зміщення свідчить про можливий ризик неправильного звітування. Перевірити цю гіпотезу ми не можемо. Ми припускаємо, що компанії, які неправильно звітують, розподілені рівномірно між кластерами та не зміщують значно центри кластерів. Варто зазначити, що за даними 2019 року, спостерігається аналогічна структура.

Також на ринку спостерігається висока концентрація однієї з груп видів страхування. Близько половини компаній мали частку премій однієї з груп видів страхування більше 60%. Це свідчить про наявність спеціалізації компаній на певному сегменті страхування. Отже, портфелі багатьох компаній можна назвати слабко диверсифікованими.

Методологія дослідження та моделі

Алгоритми кластеризації є зручним інструментом розділення спостережень на однорідні групи на основі заданих характеристик. Опис літератури наводить лише деякі з успішних випадків використання кластерного аналізу в дослідженні соціально-економічних явищ.

Важливою перевагою таких алгоритмів є нівелювання впливу судження дослідника про досліджуване явище на результати дослідження. Для дослідження бізнес-моделей страховиків ми використовуємо класичні ієрархічні, неієрархічні методи кластеризації та методи машинного навчання. Коротко опишемо застосовані методи.

Метод К-середніх є найпоширенішим неієрархічним методом. Він полягає в ітеративній мінімізації відстаней між спостереженнями в кластері, причому кількість кластерів задається на початку, тобто модель не визначає оптимальної кількості кластерів. Координати центроїдів, кількість яких відповідає кількості кластерів, на початковому етапі задаються випадково. У результаті поділ на кластери може бути нестабільним та залежати від початкових координат центроїдів.

Для уникнення цієї проблеми ми використовуємо метод початкових координат центроїдів для К-середніх, який називається “k-means++”, запропонований Arthur and Vassilvitskii (2006). Позначивши вибірку вхідних даних \mathcal{X} , а найкоротшу відстань між елементом вибірки x_i та найближчим до нього центроїдом $D(x_i)$, алгоритм можна описати покроково:

1. Визначити випадково з \mathcal{X} початковий центроїд c_1 .

2. Наступний центроїд c_j визначити з \mathcal{X} , де кожному елементу присвоєно ймовірність бути обраним:

$$p(c_j = x_i) = \frac{D(x_i)^2}{\sum_{x \in \mathcal{X}} D(x)^2}.$$

3. Повторювати крок 2, поки не буде визначено потрібної кількості центроїдів.

4. Застосувати класичний алгоритм К-середніх.

Крок 2 показує, що елементам \mathcal{X} , розташованим далі від початкового центроїду, присвоюється більша ймовірність бути обраними. Тобто центроїди розташовуються так, щоб бути відмінними один від одного. Кластери на основі процедури “k-means++” було оцінено 100 разів, з яких було обрано кластеризацію з найменшою сумою квадратів відстаней спостережень до центроїда (WCSS, Within Cluster Sum of Squares).

Підсумовуючи, центри кластерів спочатку обираються з елементів вибірки так, щоб розташовуватися далі один від одного, потім ітеративно змінюють свої координати, щоб описувати найбільшу можливу групу (кластер) елементів вибірки.

Для методу К-середніх було використано 5 кластерів, що є оптимальною кількістю кластерів з огляду на метод ліктя. Результати його застосування наведено на рисунку Б.2. Далі для всіх методів було використано розділення на п'ять кластерів.

Метод К-медоїдів, вперше описаний Kaufman and Rousseeuw (1990), є за своєю суттю дуже схожим на метод К-середніх. Ключовою характеристикою методу К-медоїдів є обрання центрів кластерів серед спостережень. Саме такі спостереження, які є “еталонними” для свого кластера, і називаються медоїдами.

За допомогою алгоритму PAM (Partitioning Around Medoids – розділення навколо медоїдів)

медоїди обираються ітеративно таким чином, щоб зменшити середню відстань від спостережень до центрів їхніх кластерів. Сучасні алгоритми методу K-медоїд пропонують швидшу оптимізацію, проте RAM залишається одним із найточніших алгоритмів вирішення цієї задачі. Тому нами було обрано саме цей метод. Початкові медоїди нами було обрано за допомогою алгоритму "k-medoids++", що є ідентичним до "k-means++" та забезпечує стійкість кластерів.

Ієрархічні методи не потребують визначення кількості кластерів перед застосуванням алгоритму. Вони будують деревоподібну структуру, що називається дендрограмою. На початку кожне спостереження формує окремий кластер. Під час подальших спостережень (кластери) на основі обраного критерію поєднуються в нові, більші кластери, поки не з'єднаються в один кластер, до якого входять усі спостереження. Для нашої цілі було обрано критерій Уорда. Кількість кластерів визначається дослідником на підставі дендрограми вже після застосування алгоритму.

За критерієм Уорда, окремий кластер поєднується з тим кластером, поєднання з яким призведе до найменшого зростання відстані між спостереженнями всередині кластера. Ця відстань, яка є аналогічною до WCSS алгоритму K-середніх, зображується на дендрограмі по вертикальній осі.

Наведемо короткий опис карти Кохонена, що є методом машинного навчання, здатним до кластеризації. Його описано в роботі Кохонена (Kohonen, 1982). Карта самоорганізації є штучною нейронною мережею, що складається з двох шарів:

1. Вхідного, на який подаються дані вибірки. Розмірність цього шару відповідає кількості характеристик, за якими відбувається поділ на групи.

2. Вихідного, що являє собою власне карту, яка складається з упорядкованих у двох (у цій праці) вимірах нейронів і має наперед визначену довільну розмірність.

Кожний нейрон карти поєднаний із кожним елементом вхідного шару вузлами, що мають ваги. Тобто кожний нейрон має набір ваг, які можуть трактуватись як опис нейрона в характеристиках, що відповідають даним вхідного шару. Алгоритм навчання карти Кохонена можна описати покроково:

1. Ваги нейронів на початку навчання є випадковими достатньо малими числами.

2. На вхідний шар подається вектор ознак x_i з \mathcal{X} та розраховується відстань (у цьому дослідженні використано квадрат евклідової відстані) між векторами x_i та w_j , де w_j – вектор ваг нейрона j вихідного шару мережі.

3. Нейрон, що є найближчим до x_i на основі кроку 2, іменується нейроном-переможцем (BMU, Best Matching Unit).

3. На підставі радіуса $\sigma(t)$ для кожного нейрона карти розраховується показник сусідства на основі функції Гауса:

$$N(t)_{BMU,j} = \exp\left(-\frac{D(BMU,j)^2}{2\sigma(t)^2}\right),$$

де $D(BMU,j)$ – топографічна відстань між нейроном-переможцем та нейроном j .

5. Ваги нейронів на карті оновлюються за такою формулою: $\Delta w_j = \alpha(t) N_j(t) (x_i - w_j)$, де $\alpha(t)$ – коефіцієнт навчання, що є спадною функцією від часу.

6. Кроки 2–5 повторюються для заданої кількості епох (циклів навчання), що визначається дослідником. Водночас потрібно звертати увагу на похибку квантування, що відображає середню відстань між вхідними даними та нейроном-переможцем, та топографічну похибку, що відображає кількість спостережень, для яких перший нейрон-переможець не є сусідом другого нейрона переможця.

У результаті навчання нейрони стають "схожими" на вхідні дані. У процесі навчання параметри $\alpha(t)$ та $\sigma(t)$ поступово зменшуються. Таким чином, чим далі заходить навчання, тим повільніше нейрони адаптують свої ваги і тим меншу "кооперацію" демонструють. Використана в цій праці спадна функція, що описує динаміку параметра $\alpha(t)$, має такий вигляд:

$$\alpha(t) = \frac{\alpha(0)}{1 + (t / (MI / 4))}, \quad (1)$$

де $\alpha(0)$ – початкове значення $\alpha(t)$, що задається дослідником;

MI – максимальна кількість епох (ітерацій), що задається дослідником;

t – порядковий номер епохи.

Ми обираємо $\alpha(0)$ для цього дослідження на рівні 0.5, $MI = 10000$. Таким чином, коефіцієнт навчання поступово зменшується від 0.5 до 0.1. Динаміка параметра $\alpha(t)$ у процесі навчання моделі є аналогічною, із початковим значенням 1. Детальніше алгоритм самоорганізаційної карти описано в роботі Honkela (1998).

Для цього дослідження з огляду на кількість спостережень використано карту розмірністю 10 x 10 (100 нейронів загалом) прямокутної топології. Поширеною є ініціалізація ваг нейронів на основі головних компонент, що спостерігаються в даних. Проте з огляду на те, що нейрони активуються (є переможцями) рівномірно на карті (рисунок Б.3) та час навчання є прийнятним, ми не застосовуємо цього методу. Динаміку топографічної помилки та помилки квантизації представлено на рисунку Б.4.

Після навчання нейрони були кластеризовані шляхом застосування методу K-середніх до їхніх ваг для порівняння результатів карти Кохонена з результатами інших методів.

Описані методи було реалізовано інструментами Python за допомогою відкритих бібліотек Scikit-learn (Pedregosa et al., 2011) та MiniSom (Vettigli, 2019).

Оцінка результатів кластеризації

Кожен із методів має свої переваги та недоліки, їхню оцінку наведено в таблиці 5.

Як бачимо, жоден із методів не можна визначити однозначно як найкращий. Тому найчастіше під час застосування кластерного аналізу обирають метод, що

Таблиця 5. Порівняння методів кластеризації

Характеристика	К-середніх	К-медоїд	Метод Уорда	Карти Кохонена
Легкість інтерпретації результатів	+	+	+	-
Наявність графічних інструментів	-	-	+	+
Стойкість до викидів	-	+-	+-	+-
Можливість застосувати оцінену модель до різних наборів даних	+	+	-	+
Легкість побудови	+	+	+	-

найкраще підходить з огляду на наявні дані та чисельні критерії якості поділу на кластери.

Для оцінки якості моделей ми використали класичний показник Калінські – Харабаша (CH-score), який оцінює якість розділення на групи на основі відстаней між спостереженнями. Також для наших цілей важливою є стійкість кластерів у часі. Бізнес-моделі відображають стійку поведінку (стратегію), а отже, для того, щоб робити висновки про бізнес-моделі та їхні ризики, важливо, щоб кластери не змінювалися суттєво в часі. Для оцінки стійкості ми застосували коефіцієнт міграції.

CH-score вперше описаний Calinski and Harabasz (1974). Також відомий як критерій співвідношення дисперсії, це відношення варіації між кластерами до варіації всередині кластерів, обидва зважені на відповідні ступені свободи. Показник обчислюється так:

$$CH - score = \frac{\left(\frac{BCSS}{k-1} \right)}{\left(\frac{WCSS}{n-k} \right)}, \quad (2)$$

де $BCSS$ – варіація між кластерами;

$WCSS$ – варіація між спостереженнями всередині кластерів;

k – кількість кластерів;

n – кількість спостережень.

Еталонного значення цього показника немає. Проте більше його значення свідчить про краще розділення на кластери. Значення CH-score є більшим тоді, коли центри кластерів перебувають далі один від одного, а спостереження в кластерах є близькими до їхніх центрів.

Якісне розділення утворює групи (кластери), що не змінюються суттєво з часом. У нашому прикладі це є принципово важливим, оскільки бізнес-модель – це стала характеристика компанії, яка не змінюється суттєво за нормальних умов функціонування. Для оцінки стабільності кластерів було розраховано коефіцієнт міграції:

$$\text{Коефіцієнт міграції} = \frac{n_m}{n_{2019-2020}}, \quad (3)$$

де n_m – кількість компаній, що на основі моделі мігрували між кластерами у 2019 та 2020 роках;

$n_{2019-2020}$ – кількість компаній, що були активними у 2019 та 2020 роках.

Міграція між кластерами відбувається в результаті дії двох факторів – зміни бізнес-моделі компанії та похибки кластеризації. Тому завелике значення коефіцієнта міграції свідчить про неточне розділення, а замале – про “перетренованість” моделі. Критичного значення цього показника немає.

Для методу Уорда було розраховано псевдокоефіцієнт міграції. Оскільки оцінену модель не можна застосувати до даних за інший рік, ми розраховували псевдокоефіцієнт міграції. Для розрахунку псевдокоефіцієнта міграції було оцінено модель на найактуальніших даних. Далі за допомогою класифікаційного методу на основі центроїдів, описаного Tibshirani et al. (2002), ми визначили кластери для даних попереднього року та застосували формулу (3).

У таблиці 6 наведено оцінки якості кластеризації для застосованих моделей.

Таблиця 6. Порівняння якості методів кластеризації

Показник	К-середніх	К-медоїд	Метод Уорда	Карта Кохонена
CH-score	68.807	68.806	77.101	-
Коефіцієнт міграції	15.8%	19.0%	20.6% (псевдо)	15.8% (між кластерами), 76.9% (між нейронами)

Очевидно, що метод Уорда відповідно до критерію CH-score дає найкращі результати кластеризації. Проте кластери є значно менш стабільними порівняно з усіма іншими методами. З огляду на цей критерій, який має велике значення з точки зору застосовуваності моделі, нами було прийнято рішення не робити висновки на основі методу Уорда. Довідково результати методу Уорда представлено на рисунку Б.6. Для зручності порівняння кластери було поійменовано аналогічно до основної моделі.

Моделі К-середніх та К-медоїд мають дуже близькі значення CH-score. Також ці значення, хоч і є нижчими за отриманий під час застосування методу Уорда, проте не надто суттєво. Модель К-середніх показує більш стабільні кластери, ніж модель К-медоїд. Також за наявності зміщень у даних модель К-медоїд може не характеризувати повністю кластери, оскільки базує висновки на одному спостереженні. Наприклад, чотири з п'яти кластерів ця модель характеризує як бізнес-моделі, що взагалі не використовують представництва у своїй діяльності. Такий висновок є помилковим з огляду на результати моделі К-середніх, що зазначені далі. Тому, зважаючи на викривлення даних, ми не базуємо свої висновки на результатах моделі К-медоїд.

Довідково результати методу К-медоїд представлено в таблиці А.2.

Для карти Кохонена було розраховано два види коефіцієнта міграції. Перший – на основі п’яти кластерів, у які згруповано нейрони. Другий – на основі всіх ста нейронів моделі. Очікувано, перший є значно меншим за другий. Цікаво, що міграція між кластерами у разі застосування карти Кохонена є майже ідентичною до методу К-середніх.

З огляду на результати оцінки якості моделей ми вирішили будувати свої висновки на результатах моделі К-середніх. Також, оскільки результати карти Кохонена є схожими з результатами К-середніх, ми використали її як інструмент візуалізації кластерів.

Далі ми побудували графік силуету (Rousseeuw, 1987) для результатів методу К-середніх, щоб більш детально оцінити результати. Коефіцієнт силуету розраховується для кожного спостереження як:

$$s_i = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad (4)$$

де $a(i)$ – середня відстань від спостереження i до членів свого кластера;

$b(i)$ – середня відстань від спостереження i до членів найближчого сусіднього кластера.

Значення коефіцієнта для моделі є середнім значенням коефіцієнта силуету всіх спостережень. Значення коефіцієнта силуету 1 свідчить про ідеальне розмежування кластерів, 0 – розмежування є випадковим, (-1) – кластери розмежовані неправильно.

-0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8

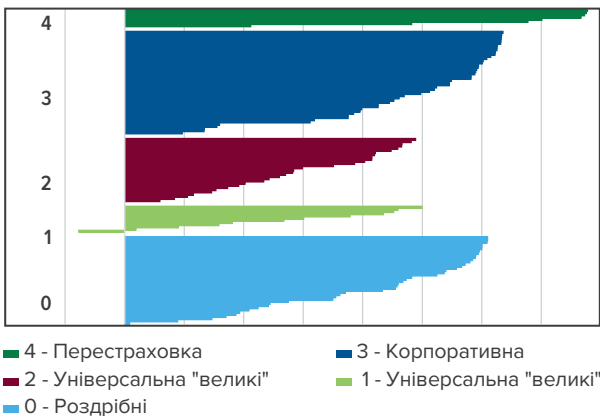


Рисунок 1. Силуети кластерів (абсциса – коефіцієнт силуету, ордината – номер кластера)

Графік коефіцієнтів силуету (рисунок 1) демонструє, що лише для одного спостереження учасники сусіднього кластера є в середньому “ближчими”, ніж учасники власного кластера. Це спостереження стосується першого кластера (універсальна “велика” модель). Він і ті, що мають значення силуету близьке до 0, можуть лежати “на межі” кластера. Загальне значення коефіцієнта (0.41) свідчить про достатньо якісне розділення груп; водночас графік показує, що “перестрахова” бізнес-модель (кластер 4) ідентифікована найкраще.

4. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Опис бізнес-моделей на основі кластерів

Модель було оцінено на основі даних 2020 року та застосовано до всіх років, наявних у вибірці (2019 і 2020 роки). Характеристики (координати центроїд) ідентифікованих кластерів зображено на рисунку 2. Координати на рисунку є стандартизованими.

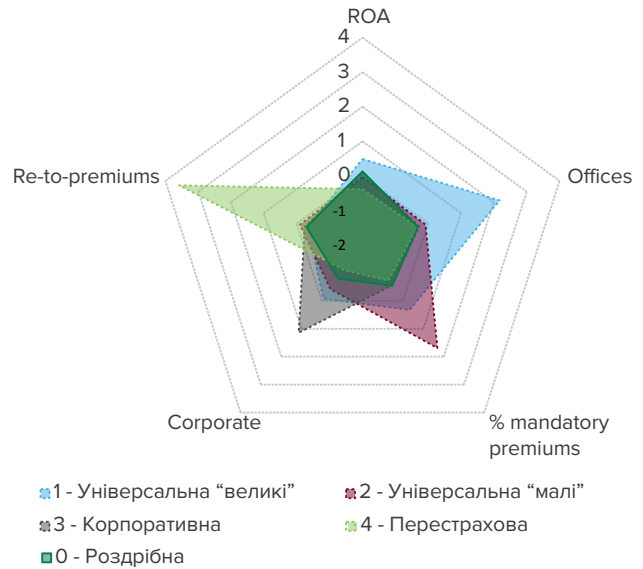


Рисунок 2. Характеристики ідентифікованих кластерів (стандартизовано)

Ідентифіковані кластери для зручності було пронумеровано та поійменовано. Ми не виключаємо, що інші однорідні групи компаній можуть бути ідентифіковані за даними та можуть викривлювати характеристики ідентифікованих нами кластерів. Проте, спираючись лише на дані та методи, описані вище, нам вдалось ідентифікувати ті бізнес-моделі, які найкраще виділяються з огляду на кількісні критерії. Далі на основі характеристик, за якими було здійснено розподіл, та описових показників наводиться короткий огляд ідентифікованих кластерів (бізнес-моделей). Таблиця А.3 підсумовує результати кластеризації у нестандартизованому вигляді.

“0 – роздрібна” бізнес-модель концентрується на страхуванні фізичних осіб (які становлять 74% у преміях) та має середній рівень рентабельності активів (3%). З огляду на невелику кількість представництв (офісів), значне відношення заробітної плати до премій компанії цієї бізнес-моделі як канал продажу використовують власних агентів, які намагаються знайти нових клієнтів. Тобто компанії “полюють” на клієнтів, а не клієнти самі до них звертаються. Страховики, що використовують цю бізнес-модель, спеціалізуються на страхуванні добровільних видів, та мають частку перестраховування у преміях у середньому на рівні 1.5%. Варто зазначити, що цей кластер має найбільший показник концентрації (76%) на одній з груп видів страхування, що може бути фактором ризику компанії цієї групи. Станом на 2020 рік у компанії кластера спостерігався найвищий показник збитковості та помірна частка перестраховування в резервах, що свідчить про вразливість компанії до андерайтингового ризику. До цього кластера у 2020 році увійшло 40 компаній, що становить 19% ринку за преміями.

Модель «1 – універсальна “великі”» – обслуговує як юридичних, так і фізичних осіб, має розподіл між обов’язковим та добровільним страхуванням 28% та 72% відповідно. Рентабельність активів є найвищою серед усіх виділених кластерів (6.5%). Характерною ознакою кластера є широке використання власних офісів (мають у середньому близько 62 представництва). Ознакою того, що компанії цієї бізнес-моделі активно використовують як власні представництва, так і агентів як канали продажу, є висока частка заробітних плат у преміях (6%). Отже, компанії намагаються диверсифікувати способи залучення клієнтів. Також компанії цього кластера мають другу найбільшу частку ризику, що перестраховується (25%), що свідчить про менші андеррайтингові ризики, а також про те, що компаніям властивий ризик дефолту контрагента (перестраховика). 12 компаній у 2020 році використовували цю бізнес-модель і займали найбільшу частку за валовими преміями – 35.5%. Варто зазначити, що з огляду на високу частку ринку в поєднанні з найнижчим показником середньої премії компанії орієнтуються на продаж великої кількості дешевих полісів.

«2 – універсальна “малі”» описується рівномірним (порівняно з іншими моделями) розподілом обов’язкового і добровільного видів страхування та страховальників фізичних і юридичних осіб у преміях (64%/36% та 63%/37% відповідно). Водночас частка премій з обов’язкових видів страхування в цій бізнес-моделі є найбільшою серед усіх кластерів. Премії з ОСЦПВ становлять 71% премій з обов’язкових видів для цього кластера. Також кластер має другий найнижчий серед усіх груп показник рентабельності активів, компанії володіють в середньому шістьма представництвами (офісами) та мають другу найвищу частку перестраховування в преміях (5.6%). Крім орієнтації на обов’язкові види, ця модель відрізняється від універсальної “великі” значним показником середньої премії, що може свідчити про те, що компанії намагаються страхувати більш дорогі ризики. До цього кластера у 2020 році увійшли 29 компаній, які сукупно становлять близько 16.5% ринку за преміями.

“3 – корпоративна” модель характеризується 89% часткою юридичних осіб у преміях, а також невисоким показником рентабельності активів (2.7%), малою кількістю представництв (офісів) та близькою до 0 часткою обов’язкових видів страхування. Виділяється значна як для компаній, що не використовують перестрахову бізнес-модель, частка вхідного перестраховування в преміях (27%). За порівняно високого рівня середньої премії (254 тис. грн) компанії цього кластера мають досить низький показник збитковості порівняно з іншими бізнес-моделями (22.5%). Передбачуваною є висока частка перестраховика в резервах, адже часто таким страховикам потрібно розділити велику експозицію корпоративного клієнта. Проте це створює компаніям кластера ризик дефолту контрагента (перестраховика). У цей кластер увійшла найбільша кількість компаній – 47, що за преміями 2020 року сукупно становлять 19% ринку.

Компанії бізнес-моделі “4 – перестрахова” мають частку перестраховування в преміях у середньому близько 81%. Рентабельність активів у компаній цього кластера в середньому є від’ємною, середня премія – більше 304 тис. грн. Компанії цього кластера взагалі не використовують представництва як канал

продажу та мають найнижчу частку заробітних плат у преміях. Частка добровільного страхування у преміях наближається до 100%. Перестраховики самі є слабо перестрахованими, що може свідчити про потенційну вразливість до андеррайтингового ризику, який вони не розділяють (не диверсифікують) між собою. Проте низьке порівняно з іншими бізнес-моделями значення показника збитковості може свідчити про незначну реалізацію андеррайтингового ризику. До кластера увійшли 8 компаній за даними 2020 року (10% ринку за преміями).

Гістограми ознак для отриманих кластерів зображено на рисунку В.6.

Нейрони на самоорганізаційній карті Кохонена в процесі тренування починають за своєю вагою бути “схожими” на вхідні дані, тобто відтворювати кластери. Карта самоорганізації Кохонена надає зручний інструмент візуалізації подібності між різними спостереженнями та характеристик цих спостережень. За допомогою карти можна побачити спостереження, що лежать на межі кластерів, та наскільки вони віддалені від інших елементів кластера.

На картах ознак на рисунку 3 зображено ваги нейронів карти, що відповідають ознакам даних (координати стандартизовані). Їх варто інтерпретувати так: компанії, для яких нейрон з координатами (1; 10) (лівий верхній кут) після навчання є переможцем (найближчим нейроном), мають середню по ринку частку премій з обов’язкових видів страхування, багато представництв та майже відсутнє перестраховування у преміях.

Сусідні нейрони є доволі схожими через механізм “кооперації” під час навчання, тож їх можна об’єднати в групи. Для цього використано алгоритм К-середніх з кількістю кластерів, що дорівнює 5, щоб інтерпретувати результати подібним чином як результати поділу методом К-середніх. Ми не наводимо центроїди цих кластерів, оскільки групування методом К-середніх було здійснено лише для того, щоб математично оцінити межі кластерів на карті Кохонена і такі центроїди повторювали би за своєю суттю центроїди попередньої моделі.

На рисунку 4 наведено результати об’єднання нейронів у кластери. Точки на рисунку позначають компанії, для яких конкретний нейрон після навчання є нейроном-переможцем. З огляду на об’єднання нейронів та карти ознак можна виділити бізнес-моделі, що є аналогічними ідентифікованим методом К-середніх.

Проте карта дає змогу побачити відстань (подібність) між спостереженнями. Топографічну відстань можна побачити відразу на карті – сусідні нейрони є схожими. Евклідову відстань між нейронами після навчання можна побачити в додатку 10. Компанії, для яких нейрон-переможець розташований на топографічній межі кластера, є “слабкими” представниками кластера та можуть змінювати свій кластер із часом. Саме про такі компанії йдеться в частині валідації результатів моделі К-середніх.

Коефіцієнт міграції для карти Кохонена між п’ятьма кластерами становить 16%, між 100 нейронами – 77%. Це свідчить про те, що міграція всередині кластерів є більшою, ніж між кластерами. Проте з огляду на цей показник висновки щодо бізнес-моделей наводяться з урахуванням результатів кластеризації методом

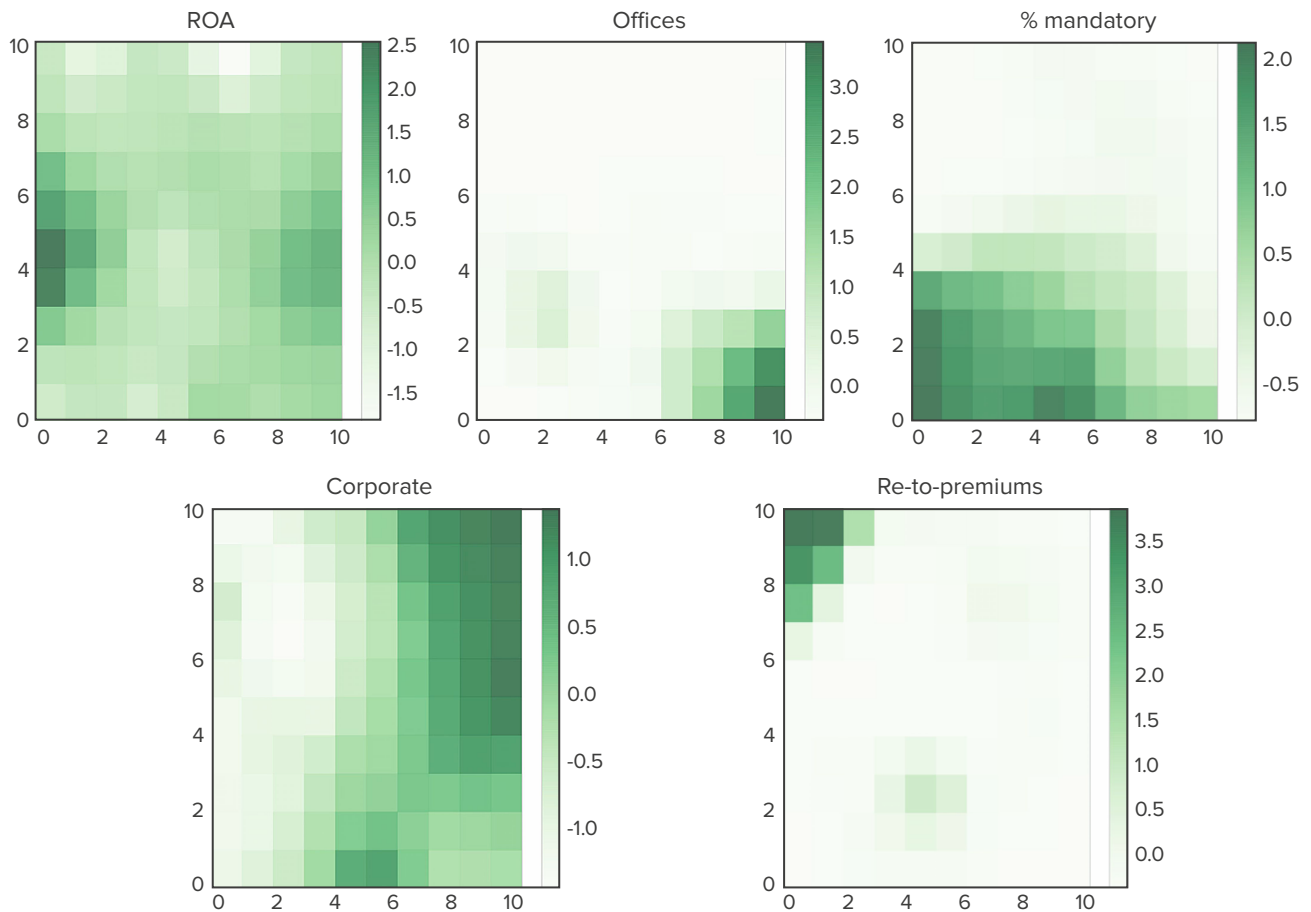


Рисунок 3. Карты характеристик Кохонена

K-средних. Але слід зазначити, що економічна суть моделей, визначених за допомогою карти Кохонена, збігається з результатами методу K-средних.

Аналіз міграції між кластерами

Маючи виділені кластери (бізнес-моделі), можна простежити динаміку складу кластерів упродовж досліджуваного періоду. Нами розраховано коефіцієнти міграції з кожного кластера в період з 2019 до 2020 року.

Як описано вище, міграція між кластерами може відбуватися під впливом двох факторів – похибки моделі та зміни компанією бізнес-моделі. Розуміючи, що кластеризація не є точним методом, ми вважали міграцію між кластерами значущою, якщо з кластера мігрувало більше 10% компаній, що до нього входять. Застосування такого порога показує значущими міграції щонайменше двох компаній з кластера та здебільшого міграції, що є більшими за значення коефіцієнта міграції моделі (15%).

Оскільки компанії, що заробили за рік премій менше ніж 5 млн грн, не було включено до алгоритму K-средних та віднесено до штучного кластера “неактивні”, є міграція з виділених бізнес-моделей до штучного кластера “неактивні”.

На рисунку 5 показано значущі міграції компаній між бізнес-моделями. Бачимо, що суттєва міграція у 2020 році відбувалася з моделі універсальна “великі” до моделі універсальна “малі”. Це очікувано, оскільки відмінність між кластерами та бізнес-моделями є незначною. Менш очевидною є міграція компаній з моделі перестрахова до моделі корпоративна. Проте, якщо подивитися на центри кластерів, можна побачити, що корпоративна модель є ближчою до перестрахової, ніж інші, адже в компанії перестрахової моделі є невелика частка премій від юридичних осіб, а в компанії моделі корпоративна – частка премій із вхідного перестраховання.

З огляду на міграції до групи “неактивні” компанії, що використовують бізнес-моделі корпоративна, роздрібна та перестрахова, можуть мати більший ризик виходу з ринку, а отже, і розглядаються менш стійкими. Більше

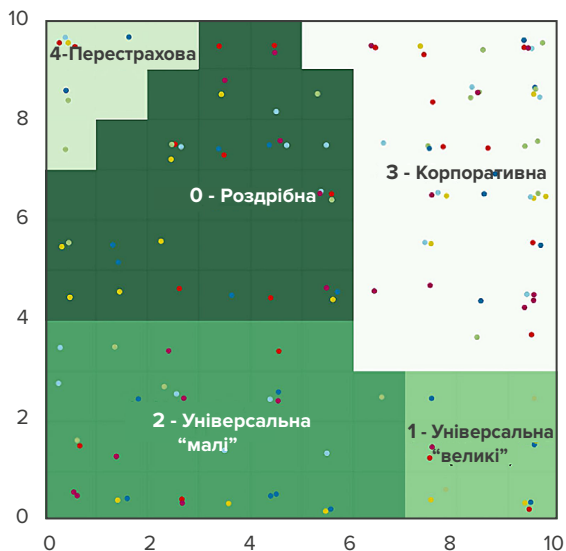


Рисунок 4. Групування кластерів на карті Кохонена

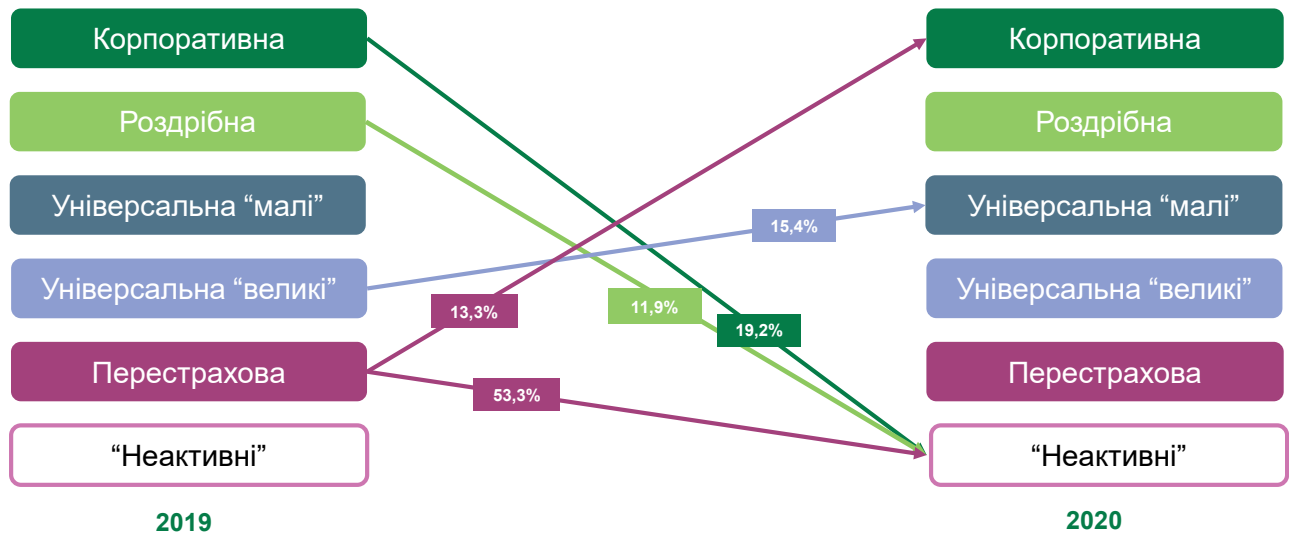


Рисунок 5. Міграції між бізнес-моделями

половини компаній моделі перестрахова вийшли з ринку у 2020 році, що може слугувати сигналом для регулятора, що пильний нагляд є необхідним.

Описуючи бізнес-моделі, ми зауважили, що корпоративні страховики широко використовують вихідне перестраховання своїх ризиків, тобто у своїй роботі залежать від перестраховиків. Отже, логічним є високий рівень одночасного виходу з ринку компаній саме цих двох кластерів.

Можна оцінити робастність висновків, що ґрунтуються на міграції до групи "неактивні". До цієї групи ми відносили компанії, обсяг премій яких за рік був меншим 5 млн грн, отже, мігрувати могли дуже малі компанії, для яких премії

близько 5 млн грн є нормою через варіацію обсягу премій рік від року. Менше 10 млн грн премій у 2019 році мали лише 24% компаній, що мігрували в "неактивні". Медіанне значення обсягу премій серед компаній, що мігрували до групи, становить 175 млн грн, середнє – 397 млн грн. Оскільки стрімке скорочення обсягу премій із таких значень свідчить про кризу в діяльності компанії, результати можна вважати робастними. Лише 16% цих компаній мали ненульовий обсяг премій у 2020 році. Тобто більшість повністю припинила страхову діяльність.

Більш детально міграції між кластерами та їхні причини можна оцінити за допомогою карти Кохонена. На рисунку 6 зображені компанії, які або у 2019 році, або у 2020 році входили до відповідних кластерів. Варто звернути увагу,

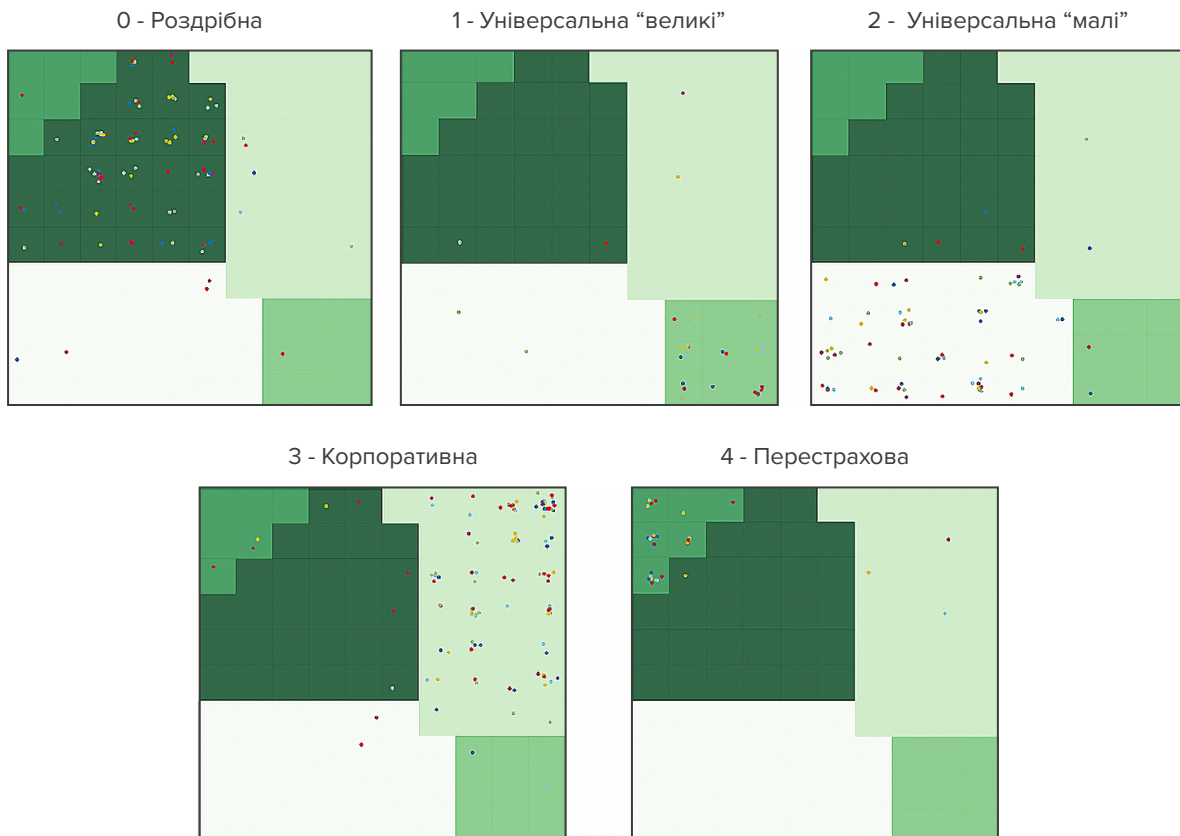


Рисунок 6. Міграції на карті Кохонена

що якщо на рисунку 5 показано односторонні міграції, то на карті Кохонена зображено двосторонні (разом компанії, які мігрували з кластера в кластер).

На рисунку можна спостерігати, що міграція відбувається здебільшого між сусідніми кластерами та нейронами. Коефіцієнт міграції між кластерами для карти Кохонена є близьким до моделі К-середніх, він становить 15.8%.

Наприклад, показано що значна міграція між кластером перестрахова та кластером корпоративна, яку ідентифіковано за обома моделями, відбувається за рахунок різкого згортання перестраховиками своєї діяльності та початку обслуговування напряму корпоративних клієнтів. Оскільки ця міграція відбувається не між сусідніми нейронами та кластерами, можна стверджувати, що компанії не перебували на межі кластерів, а дійсно суттєво змінили бізнес-модель.

На відміну від моделі К-середніх, на карті Кохонена ми не спостерігаємо значної міграції всередині бізнес-моделі універсальна. Обидва кластери цієї моделі демонструють незначну міграцію компаній до та з інших кластерів, яку ми не можемо назвати стійкою. Цікавим є те, що для малих компаній цієї бізнес-моделі міграція відбувається здебільшого із сусідніми нейронами на межі кластера, для великих – міграція тільки далеко від межі кластера.

Зобразивши компанії, які припинили здійснення страхової діяльності, на карті Кохонена, можна виділити її зони, що характеризуються високим ризиком (рисунок 7).

Результати карти Кохонена є консистентними з моделлю К-середніх; найбезпечнішою бізнес-моделлю можна вважати модель універсальна. Компанії, що використовують моделі перестрахова та корпоративна, емпірично є найменш стійкими.

Особливо ризиковою зоною моделі корпоративна є правий верхній кут карти. Там розміщуються компанії, у яких частка юридичних осіб у преміях близька до 100% та які здійснюють види добровільного страхування. Варто зазначити, що збитковість навряд

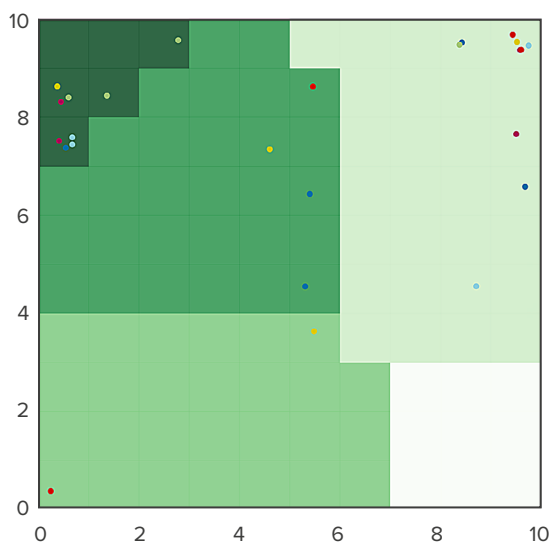


Рисунок 7. Компанії, що мігрували до групи "неактивні"

чи є причиною ризиковості цих компаній, адже вони мають близький до середньоринкового рівень рентабельності активів.

Для моделі перестрахова ризиковою є нижня частина кластера на карті. Це компанії, які займаються як перестрахованням, так і прямим страхуванням добровільних видів. Можна зробити висновок, що більш стабільні перестраховики займаються виключно перестраховою діяльністю.

Можна помітити, що регіон карти, який характеризується найвищою рентабельністю активів, не показує абсолютно міграцій до групи "неактивні". Тобто прибуткова діяльність підвищує стабільність компаній.

5. ВИСНОВКИ

У цій статті ми досліджуємо ринок страхування іншого, ніж страхування життя в Україні. Зокрема, ми намагаємося знайти стійкі та економічно обґрунтовані способи ведення бізнесу, які використовують компанії (бізнес-моделі). Бізнес-модель висвітлює не тільки операційні, а й ризикові характеристики компанії. Таким чином, знання бізнес-моделей і вміння визначити, яку модель використовує конкретна компанія, має велике значення в процесі нагляду.

Спершу ми з'ясуємо, які кількісні показники можуть допомогти описати бізнес-модель страхової компанії. Потім застосуємо набір методів кластеризації до показників на рівні компанії, розрахованих на основі нормативної бази даних, і групуємо компанії в кластери. Якщо ми знаємо, що ці групи стабільні в часі і формуються на підставі показників, які описують бізнес-модель, то опис кластера сам собою є описом бізнес-моделі. Використання чітко визначених алгоритмів і показників продуктивності дає змогу значною мірою виключити особисті судження. Нарешті, оскільки ми розділили компанії на кластери, ми можемо визначити, як компанії змінювали свої кластери протягом періоду дослідження.

Ми застосуємо набір алгоритмів кластеризації до обраних даних. Зокрема, ми виконуємо кластеризацію за допомогою ієрархічного методу Уорда, К-середніх, К-медоїдів і карти Кохонена. Ми виявили, що метод К-середніх забезпечує найкраще поєднання якості поділу кластерів та їх стабільності протягом тривалого часу. Ми також використовуємо карту Кохонена як зручний інструмент візуалізації результатів кластеризації, оскільки кластери карти Кохонена мають таке саме економічне значення, як і кластери К-середніх.

На основі кількісних даних ми виділяємо чотири різних бізнес-моделі страховиків на ринку України: роздрібна, корпоративна, універсальна (поділена на два кластери), перестрахова. Шостий кластер є штучно виділеним, до нього входять страхові компанії, валові премії яких за рік становили менше 5 млн грн та які для цілей цього дослідження вважалися неактивними. Дослідження також описує виділені на підставі ключових кількісних показників бізнес-моделі, що їх характеризують.

Компанії роздрібної бізнес-моделі страхують фізичних осіб та схильні концентруватися на певній групі видів страхування. Така концентрація та низький рівень вихідного перестраховання робить їх вразливими до андерайтингового ризику.

Великі універсальні страховики – це здебільшого відомі страхові компанії, які користуються довірою споживачів, мають багато офісів та високу рентабельність. Вони орієнтуються на продаж великої кількості дешевих полісів.

Малі універсальні страховики схильні страхувати обов'язкові види, зокрема ОСЦПВ. Таким чином, ризики цієї бізнес-моделі тісно пов'язані з ризиками автоцивільної відповідальності. Ці компанії зазвичай мають низьку рентабельність.

Корпоративні страховики концентруються на юридичних особах та страхують дорогі ризики. Вони широко використовують вихідне перестраховання для зменшення андерайтингового ризику. Проте це робить їх вразливими до ризику дефолту контрагента.

Ми вважаємо перестраховиків найменш рентабельними на ринку, вони перестраховують здебільшого добровільні види страхування. Ми виявили, що перестраховики самі є слабко перестрахованими, що робить їх вразливими до андерайтингового ризику.

Дослідження також показує міграцію страхових компаній між кластерами. Відповідно до моделі компанії,

що використовують корпоративну та перестрахову бізнес-моделі з 2019 до 2020 років найчастіше виходили з ринку, що може свідчити про те, що такі компанії потребують більшої уваги регулятора. Водночас роздрібна та універсальна моделі є найбільш стабільними, а отже, можуть вважатися найменш ризиковими. Тож запропоноване поєднання методів можна вважати ефективним для цілей нагляду за ринком.

Ця стаття створює підґрунтя для подальших досліджень у двох сферах. По-перше, ми вважаємо ідентифіковані в цьому дослідженні кластери доволі широкими, хоча вони і відповідають ключовим напрямкам діяльності компаній. Отже, виділення бізнес-моделей, що є більш вузько орієнтованими, на основі ідентифікованих у цій праці кластерів було би логічним продовженням розроблення теми. По-друге, з огляду на описану емпіричну залежність рівня ризику страховика від бізнес-моделі, яку він використовує, вкрай важливим є дослідження факторів ризику, що діють на компанії різних кластерів. Ключовим чинником, що сприяв би подальшому розробленню цієї тематики, ми бачимо доступність детальних і надійних даних щодо компаній на страховому ринку України.

ЛІТЕРАТУРА

- Abbas, S. A., Aslam, A., Rehman, A. U., Abbasi, W. A., Arif, S., Kazmi, S. Z. H. (2020). K-Means and K-Medoids: Cluster analysis on birth data collected in city Muzaffarabad, Kashmir. *IEEE Access*, 8, 151847-151855. <https://doi.org/10.1109/ACCESS.2020.3014021>
- Abolmakarem, S., Abdi, F., Khalili-Damghani, K. (2016). Insurance customer segmentation using clustering approach. *International Journal of Knowledge Engineering and Data Mining*, 4(1), <https://doi.org/10.1504/IJKEDM.2016.082072>
- Ahmar, A. S., Napitupulu, D., Rahim, R., Hidayat, R., Sonatha, Y., Azmi, M. (2018). Using K-Means clustering to cluster provinces in Indonesia. *Journal of Physics: Conference Series*, 1028, 012006. <http://doi.org/10.1088/1742-6596/1028/1/012006>
- Arthur, D., Vassilvitskii, S. (2007). k-means++: the advantages of careful seeding. *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (p./pp. 1027--1035), Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.
- Bach, M. P., Vlahović, N., Pivar, J. (2020). Fraud prevention in the leasing industry using the Kohonen self-organising maps. *Organizacija*, 53(2), 128-145. <https://doi.org/10.2478/orga-2020-0009>
- Caliński, T., Harabasz, J. (1972). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1-27. <https://doi.org/10.1080/03610927408827101>
- Honkela, T. (1998). Description of Kohonen's self-organizing map. In Honkela, T. *Self-Organizing Maps in Natural Language Processing*. Helsinki: Helsinki University of Technology. Retrieved from <http://www.mlab.uiah.fi/~timo/som/thesis-som.html>
- Kaufman, L., Rousseeuw, P. J. (1990). Partitioning around medoids (Program PAM). In Kaufman, L., Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis* Finding Groups in Data, pp. 68-125. John Wiley & Sons. <https://doi.org/10.1002/9780470316801.ch2>
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59–69. <https://doi.org/10.1007/BF00337288>
- Kramarić, T. P., Bach, M. P., Dumičić, K., Žmuk, B., Žaja, M. M. (2017). Exploratory study of insurance companies in selected post-transition countries: Non-hierarchical cluster analysis. *Central European Journal of Operations Research*, 26(3), 783–807. <https://doi.org/10.1007/s10100-017-0514-7>
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830. Retrieved from <http://jmlr.org/papers/v12/pedregosa11a.html>
- Rashkovan, V., Pokidin, D. (2016). Ukrainian banks' business models clustering: Application of Kohonen neural networks. *Visnyk of the National Bank of Ukraine*, 238, 13-38. <https://doi.org/10.26531/vnbu2016.238.013>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Tibshirani, R., Hastie, T., Narasimhan, B., Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 99(10), 6567-6572. <https://doi.org/10.1073/pnas.082099299>
- Velykoivanenko, H., Beschastna, G. (2018). Analysis of the stability and rating of Ukrainian insurance companies. *Modelling and Information Systems in Economics*, 95, 65-81. Retrieved from <https://ir.kneu.edu.ua:443/handle/2010/30980>
- Vettigli, G. (2019). MiniSom: Minimalistic and Numpy-Based Implementation of the Self Organizing Map (release 2.1.5). 2019. Retrieved from <https://github.com/JustGlwing/minisom>
- Wang X., Keogh E. (2008) A clustering analysis for target group identification by Locality in motor insurance industry. *Soft Computing Applications in Business. Studies in Fuzziness and Soft Computing*, 230, 113–127. https://doi.org/10.1007/978-3-540-79005-1_7
- Zaqueu, J. R. (2019). Customer Clustering in the Health Insurance Industry by Means of Unsupervised Machine Learning: An Internship Report. University of Lisbon, Information Management School. Retrieved from <https://run.unl.pt/bitstream/10362/89468/1/TAA0043.pdf>

ДОДАТОК А. ТАБЛИЦІ

Таблиця А.1. Описова статистика змінних у 2019 році

а) описова статистика змінних моделі

	ROA	Offices	Re-to-premiums	% of mandatory premiums	Corporate
Середнє	0.10	6.26	0.17	0.49	0.13
Ст. відхилення	0.25	18.44	0.25	0.34	0.27
Мін.	-0.22	0.00	0.00	0.00	0.00
Q (25%)	0.02	0.00	0.00	0.18	0.00
Q (50%)	0.05	0.00	0.03	0.44	0.01
Q (75%)	0.09	0.00	0.29	0.83	0.08
Макс.	2.19	115.00	0.90	1.00	1.00

б) описова інформативних показників компаній

	Re-to-provisions	Loss ratio	Mean premium	Wages/Premiums	Concentration
Середнє	0.23	0.49	243.71	0.02	0.70
Ст. відхилення	0.25	0.50	1,199.14	0.02	0.18
Мін.	0.00	-0.63	0.00	0.00	0.33
Q (25%)	0.03	0.09	0.79	0.00	0.57
Q (50%)	0.14	0.43	2.69	0.01	0.69
Q (75%)	0.37	0.73	26.32	0.02	0.81
Макс.	1.72	3.33	13,768.22	0.07	1.00

Таблиця А.2. Результати методу К-медоїд

а) координати центрів кластерів

	ROA	Offices	% of mandatory premiums	Corporate	Re-to-premiums
0 – роздрібна	0.002	0.000	0.000	0.242	0.004
1 – універсальна "великі"	0.067	44.000	0.242	0.433	0.017
2 – універсальна "малі"	0.001	0.000	0.663	0.342	0.068
3 – корпоративна	0.011	0.000	0.001	0.913	0.004
4 – перестрахова	0.001	0.000	0.000	0.038	0.893

б) кількість компаній, що увійшли до кластерів у досліджуваний період

	2019	2020
0 – роздрібна	44	40
1 – універсальна "великі"	14	14
2 – універсальна "малі"	28	28
3 – корпоративна	52	46
4 – перестрахова	14	8

Таблиця А.3. Опис кластерів

а) координати центрів кластерів

	ROA	Offices	% of mandatory premiums	Corporate	Re-to-premiums
0 – роздрібна	0.030	0.350	0.050	0.242	0.015
1 – універсальна "великі"	0.065	62.330	0.277	0.482	0.008
2 – універсальна "малі"	0.014	6.140	0.637	0.368	0.057
3 – корпоративна	0.027	1.190	0.047	0.887	0.027
4 – перестрахова	-0.020	0.000	0.001	0.141	0.810

б) додаткові описові характеристики кластерів (2020 р.)

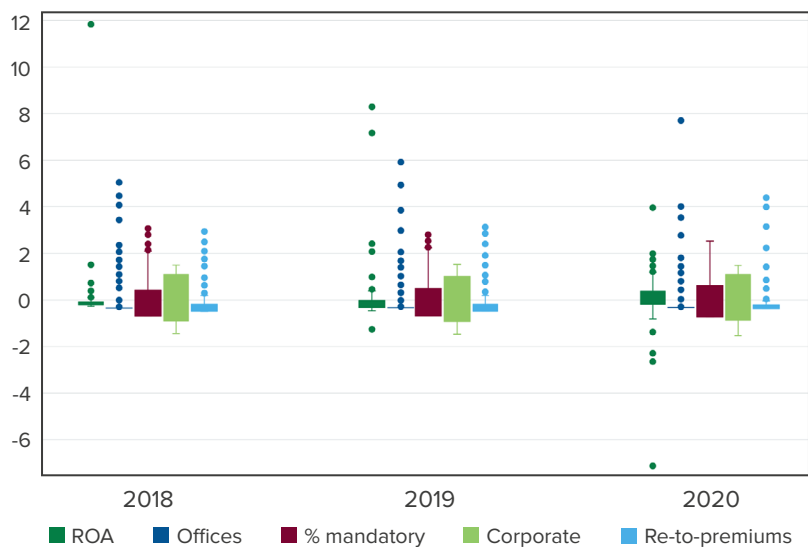
	Re-to-provisions	Loss ratio	Mean premium	Wages/Premiums	Concentration
0 – роздрібна	0.186	0.457	5.424	0.065	0.759
1 – універсальна "великі"	0.250	0.381	1.669	0.060	0.466
2 – універсальна "малі"	0.168	0.390	41.900	0.064	0.637
3 – корпоративна	0.275	0.225	254.20	0.051	0.701
4 – перестрахова	0.119	0.049	303.97	0.001	0.679

в) кількість компаній, що увійшли до кластерів у досліджуваний період

	2019	2020
0 – роздрібна	42	40
1 – універсальна "великі"	13	12
2 – універсальна "малі"	29	29
3 – корпоративна	52	47
4 – перестрахова	15	8

ДОДАТОК Б. РИСУНКИ

а) до коригування



б) після коригування

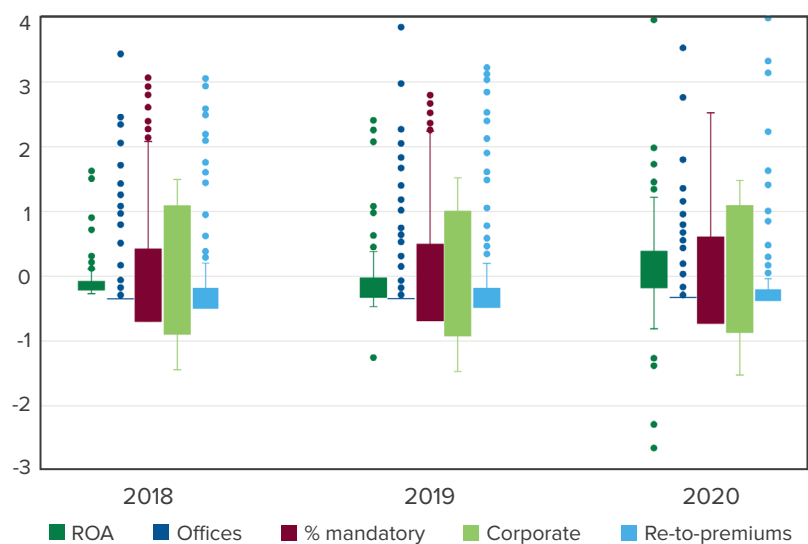


Рисунок Б.1. Розподіл значень змінних до та після коригування викидів, значення, роки

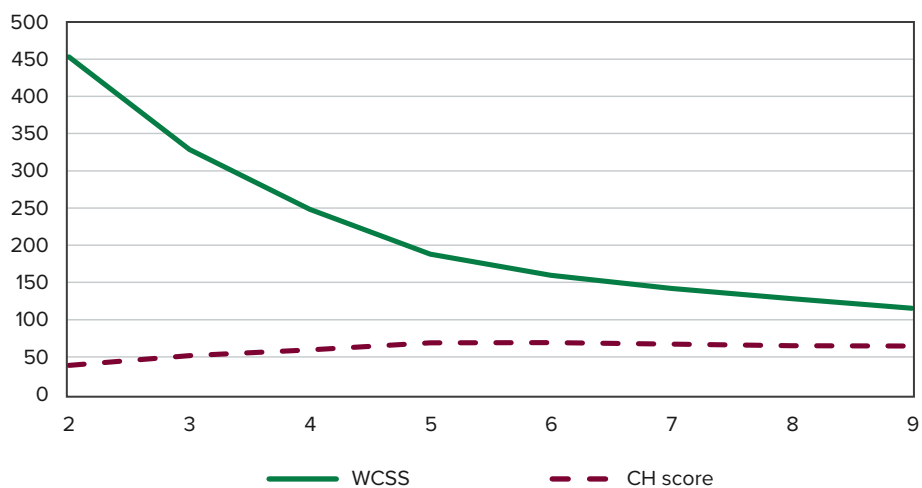


Рисунок Б.2. Критерії вибору кількості кластерів

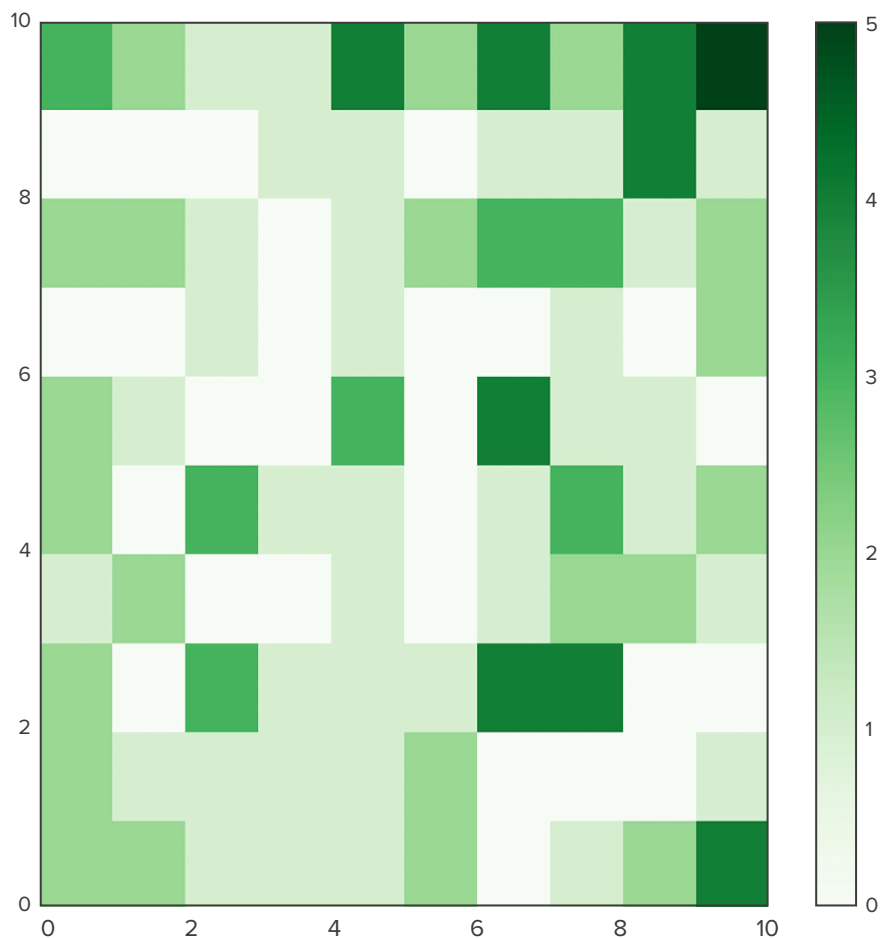


Рисунок Б.3. Кількість активацій нейронів на карті Кохонена

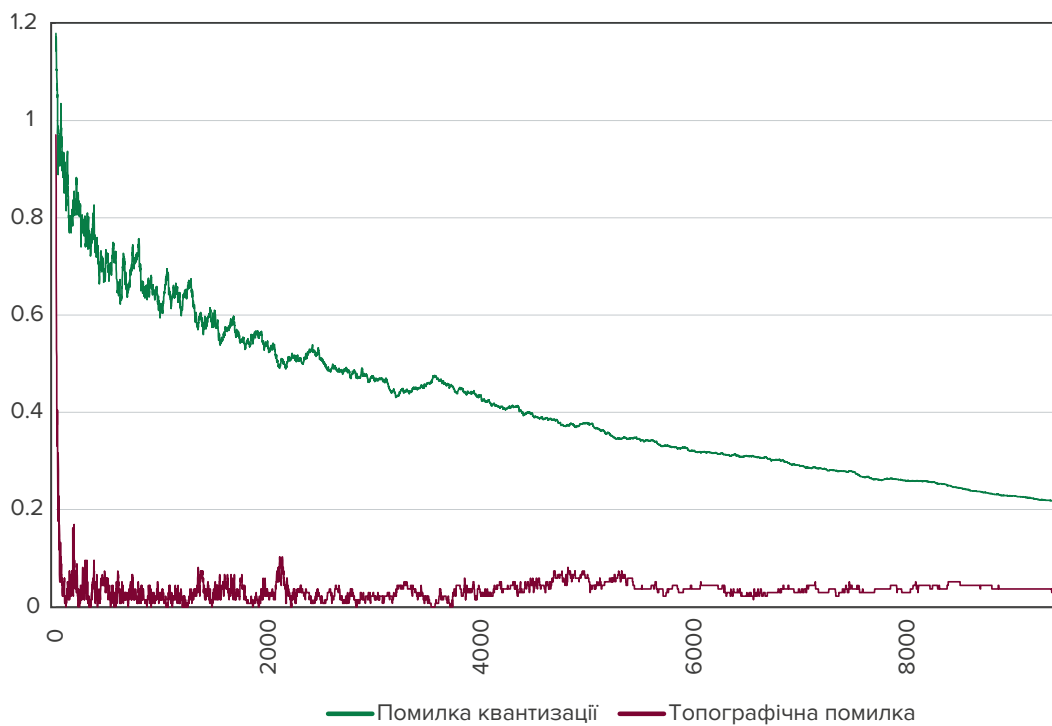
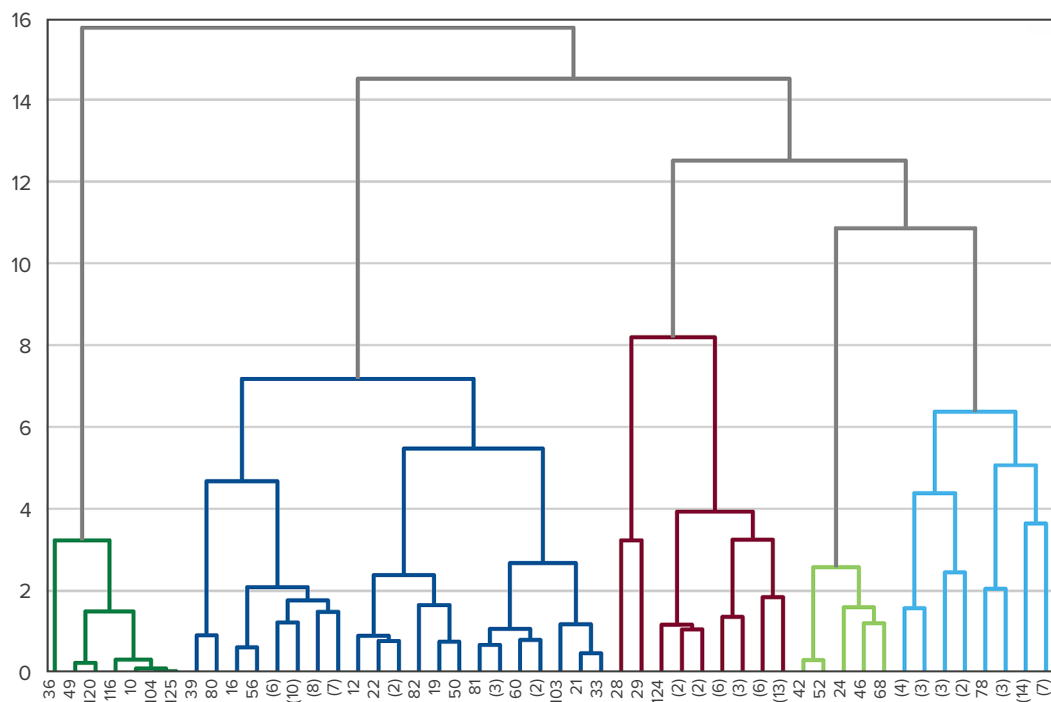


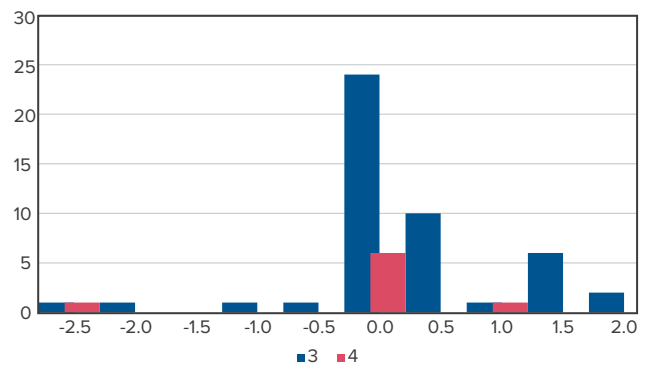
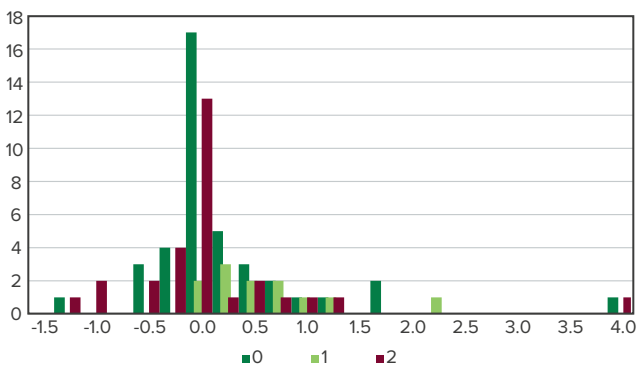
Рисунок Б.4. Динаміка навчання мережі Кохонена



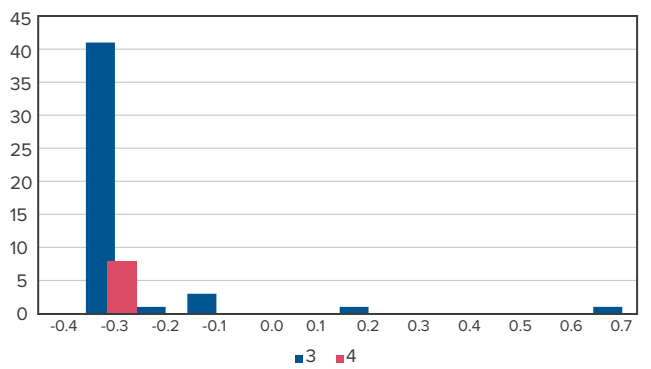
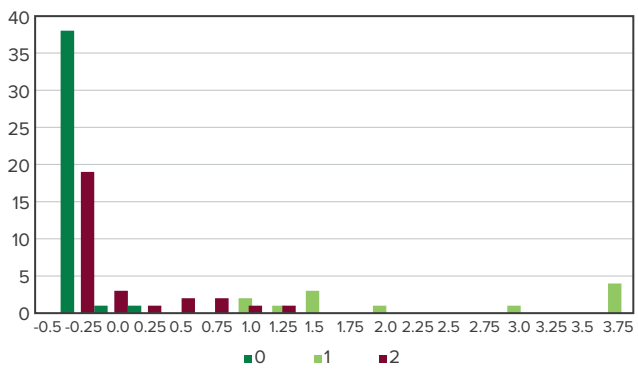
	ROA	Offices	% of mandatory premiums	Corporate	Re-to-premiums	Кількість компаній (2020)
0 - роздрібна	0.046	0.143	0.044	0.207	0.009	35
1 - універсальна "великі"	0.057	92.800	0.278	0.487	0.001	5
2 - універсальна "малі"	0.002	6.838	0.537	0.393	0.066	37
3 - корпоративна	0.034	5.269	0.062	0.847	0.026	52
4 - перестрахова	-0.032	0.000	0.001	0.096	0.849	7

Рисунок Б.5. Результати методу Уорда

a) ROA



б) Offices



в) % mandatory premiums

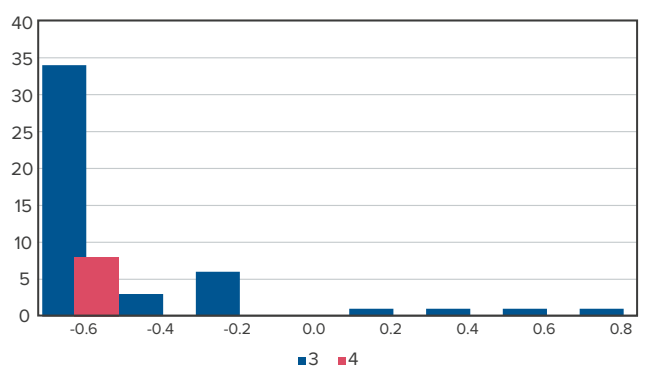
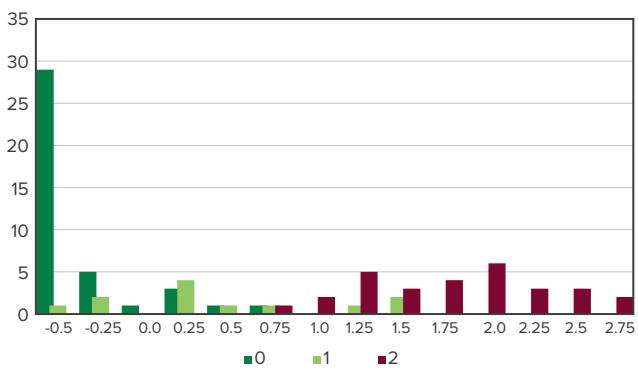
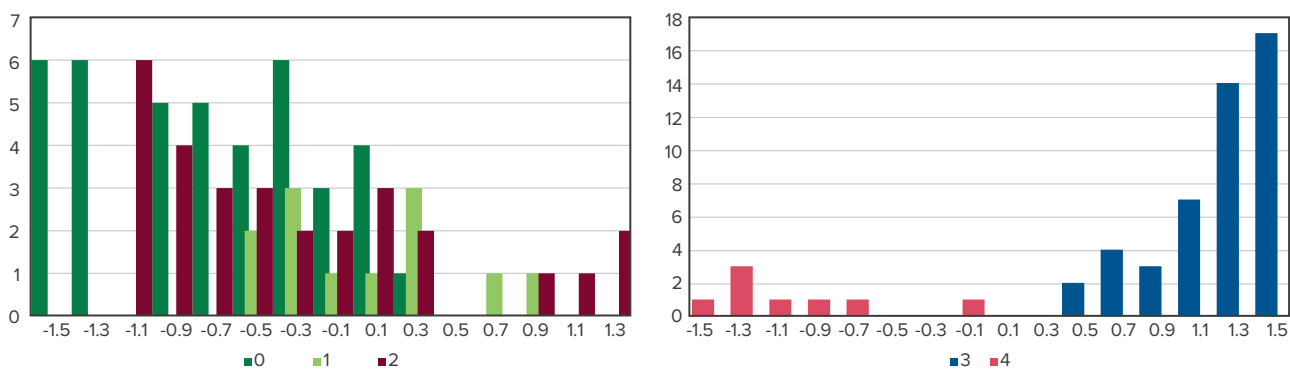


Рисунок Б.6. Гістограми ознак виділених кластерів (стандартизовано)

г) Corporate



д) Re-to-premiums

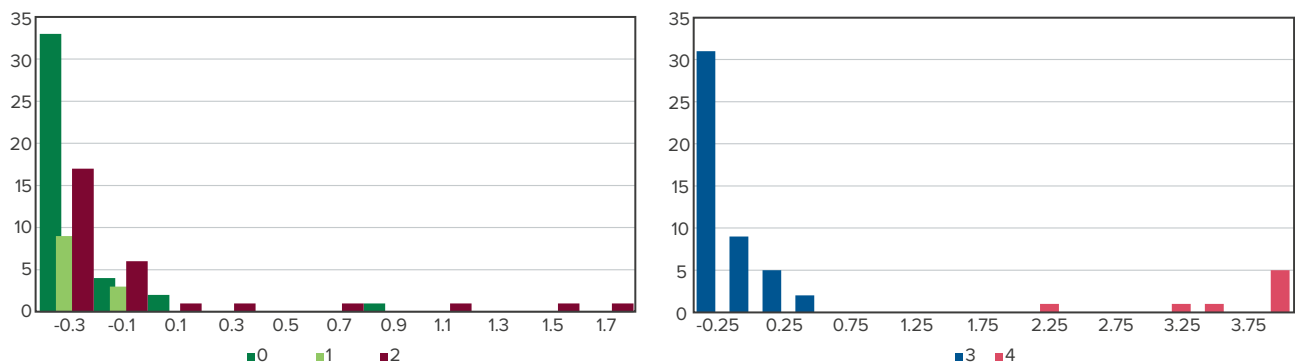


Рисунок Б.6 (продовження). Гістограми ознак виділених кластерів (стандартизовано)

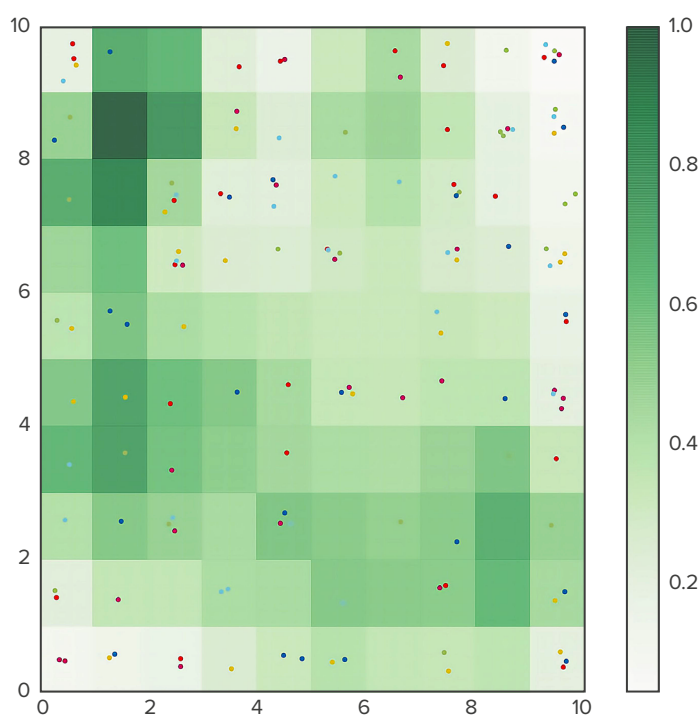


Рисунок Б.7. Евклідова відстань між нейронами (нормалізовано)