

РОЗРОБКА ДЕРЕВ РІШЕНЬ ДЛЯ ПРОГНОЗУВАННЯ ФІНАНСОВОЇ КРИЗИ В СТРАХОВИХ КОМПАНІЯХ УКРАЇНИ

У статті проаналізовано можливість застосування дерев рішень для прогнозування фінансової кризи в українських страхових компаніях. Використовуючи сформовані набори ознак, побудовано кілька моделей дерев рішень, серед яких шляхом порівняння класифікаційних характеристик за два періоди обрано найкращу з позиції чутливості та точності модель. На підставі аналізу окреслено шляхи вдосконалення практики застосування методу дерев рішень для передбачення кризових явищ у страхових компаніях України.

Ключові слова: дерева рішень, прогнозування фінансової кризи, фінансові показники, страхова компанія.

Вступ та постановка проблеми

Забезпечення фінансової стійкості підприємств є одним з основних завдань управлінців, адже саме з фінансовою стійкістю прямо пов'язані здатність будь-якої компанії протидіяти різноманітним негативним факторам та можливість реалізації довгострокових цілей власників. Прогнозування фінансових криз – важливий елемент систем антикризового менеджменту та управління фінансовою стійкістю на підприємстві, спрямований на своєчасне виявлення загрозливих симптомів у фінансовому стані та результатах діяльності компаній з метою попередження кризових явищ та забезпечення адекватного рівня стабільності фінансової системи компанії.

Страховики є фінансовими компаніями, функціонування яких має низку характерних рис у порівнянні з пересічними підприємствами. Серед них – фактична відсутність запасів, випадковий характер страхових зобов'язань, істотність неопераційних грошових потоків, особливості оподаткування тощо. Відмінності в діяльності страхових компаній обмежують застосування типових практик антикризового управління загалом та прогнозування криз зокрема. Окрім цього, український ринок також має чимало особливостей у порівнянні з ринками фінансових послуг інших країн, що накладає обмеження на використання зарубіжного досвіду в цій сфері. Таким чином, існує потреба розробки та впровадження методів прогнозування фінансової кризи, які б враховували особливості діяльності українських страховиків.

Аналіз останніх досліджень і публікацій

Проблемі прогнозування фінансової кризи на підприємствах та банкрутства компаній було приділено чимало уваги як зарубіжними, так і вітчизняними науковцями. Чимало праць присвячено застосуванню класичних статистичних моделей, таких як дискримінантний аналіз і логістична регресія (Е. Альтман, Дж. Ольсон, Т. Король, А. Шаріту, О. Терещенко, А. Матвійчук, О. Гребенікова та ін.), а також використанню методів штучного інтелекту, таких як нейронні мережі, еволюційні алгоритми, нечіткі та приблизні множини та ін. (М. Чен, Ч. Чуанг, З. Янг, А. Матвійчук, О. Шарапов та ін.). Утім, перелік досліджень, присвячених саме прогнозуванню фінансової кризи в страхових компаніях, є надзвичайно обмеженим. Більшість таких моделей були побудовані західними науковцями для реалій ринків США та Європи (П. Брокет, Н. Ш'ет, А. Клеффнер, С. Лі, І. Шарп та ін.), тоді як праці українських науковців здебільшого присвячені дослідженню суміжних тем, зокрема, було розроблено моделі управління страховиками (О. Клепікова, З. Соколовська та ін.) та моделі експрес-діагностики їхнього фінансового стану (К. Грозава та ін.).

Невирішені частини проблеми

Обмежена увага науковців до проблеми застосування економіко-математичних методів, особливо методів штучного інтелекту, для прогнозування фінансової кризи в страхових компаніях в Україні диктує потребу аналізу можливості використання дерев рішень у зазначеній сфері.

Таблиця 1. Перелік фінансових показників страхових компаній

Показник	Назва змінної
Ліквідні активи / Валові страхові премії*	$[C + CFI + R] / S$
Ліквідні активи / Валюта балансу	$[C + CFI + R] / TA$
Приріст ліквідних активів	$[C + CFI + R] Gr$
Грошові кошти / Поточні зобов'язання	C / CL
Грошові кошти / Валюта балансу	C / TA
Грошовий потік (чистий) / Борг	NCF / D
Грошовий потік від операційної діяльності / Поточні зобов'язання	CFO / CL
Грошовий потік від операційної діяльності / Валові страхові премії	CFO / S
Грошовий потік від операційної діяльності / Валюта балансу	CFO / TA
Грошовий потік від операційної діяльності / Борг	CFO / D
Грошовий потік від операційної діяльності / Зобов'язання	CFO / TL
Поточні активи / Валові страхові премії	CA / S
Поточні активи / Валюта балансу	CA / TA
Поточні зобов'язання / Валюта балансу	CL / TA
Прибуток до сплати відсотків і податків / Відсотки	$EBIT / I$
Прибуток до сплати відсотків і податків / Валюта балансу	$EBIT / TA$
Власний капітал / Валові страхові премії	E / S
Власний капітал / Валюта балансу	E / TA
Власний капітал / Зобов'язання	E / TL
Приріст власного капіталу	$E Gr$
Валові страхові премії / Власний капітал	GP / E
Ріст чистого доходу	$GP Gr$
Страхові резерви (чисті) / Власний капітал	NIR / E
Запаси / Валові страхові премії	I / S
Інвестиційні витрати / Ліквідні активи	$IE / [C + CFI + R]$
Інвестиційний прибуток / Інвестовані активи	$II / [C + CFI + LTFI]$
Ліквідні активи / Страхові резерви (чисті)	$[C + CFI + R] / IR$
Чистий грошовий потік від фінансової діяльності	CFF
Чистий грошовий потік від інвестиційної діяльності	CFI
Чистий грошовий потік від операційної діяльності	CFO
Валюта балансу (натуральний логарифм)	$\ln(TA)$
Довгостроковий борг / Валюта балансу	LTD / TA
Чистий прибуток / Власний капітал	NI / E
Чистий прибуток / Валові страхові премії	NI / S
Чистий прибуток / Валюта балансу	NI / TA
Операційний прибуток / Валові страхові премії	NOI / GP
Валові страхові премії / Валюта балансу	S / TA
Операційні витрати / Операційний прибуток	OE / OI
Операційний прибуток / Валюта балансу	OI / TA
Коефіцієнт швидкої ліквідності	QR
Нерозподілений прибуток (непокриті збитки) / Валюта балансу	RE / TA
Борг / Власний капітал	TD / E
Борг / Валюта балансу	TD / TA
Зобов'язання / Валюта балансу	TL / TA
Власні оборотні кошти / Власний капітал	WC / E
Власні оборотні кошти / Валові страхові премії	WC / S
Власні оборотні кошти / Валюта балансу	WC / TA
Рівень виплат	PR

* Надалі, якщо чисельник є показником потоку, а знаменник – запасу (і навпаки), береться середнє значення показника запасу для відповідності одиниць виміру.

Метою дослідження є розробка моделі дерев рішень на основі звітних даних українських страховиків, яка може бути використана для прогнозування фінансової кризи з прийнятною точністю.

Дані та метод

Інформаційною базою для побудови моделей дерев рішень слугували звітні дані українських компаній із ризикового страхування за 2010 (308 компаній) та 2011 (314 компаній) роки, зібрані з різних джерел, серед яких офіційний веб-сайт Агентства з розвитку інфраструктури фондового ринку України, публічна база даних Національної комісії з регулювання ринків фінансових послуг України, номери друкованого видання «Україна Бізнес Ревю» та корпоративні веб-сторінки страховиків. Статус страховиків (функціонує, або «0», та банкрут, або «1») визначався з використанням Комплексної інформаційної системи Національної комісії з регулювання ринків фінансових послуг України.

Перелік із 48 фінансових індикаторів для використання в процесі моделювання було сформовано на основі аналізу наукових праць, присвячених прогнозуванню криз у страхових компаніях [1–3; 5; 7], а також досліджень, у яких узагальнюється практика побудови моделей прогнозування банкрутства для підприємств різних галузей [6]. До переліку не було включено індикатори, які неможливо розрахувати на основі наявної звітної інформації. У таблиці 1 наведено перелік фінансових показників та відповідні назви змінних латиницею.

Після проведення математико-статистичного аналізу даних було виключено індикатори з великим числом пропусків та сформовано набори некорельованих показників (набори № 1–4). Окрім цього, застосувавши алгоритм вибору ознак, а саме метод мінімальної надлишковості – максимальної доречності з додатковим урахуванням зміни важливості ознак між роками, було визначено перелік

найбільш релевантних індикаторів страховиків для процесу класифікації (набори № 5–7). При цьому використовувалися оброблені вибірки.

Для побудови моделей прогнозування фінансової кризи в страхових компаніях було використано дерева рішень. Цей метод на противагу статистичним методам не має критичних обмежень застосування та на відміну від моделей «чорного ящика» відкрито представляє логіку класифікації. Логічні правила, які одержуються в результаті побудови та застосування дерев рішень, є легшими для інтерпретації, що дає змогу більш широко застосовувати цей метод у процесі прийняття рішень [4, с. 261]. Дослідниками було розроблено багато алгоритмів дерев рішень, серед яких C4.5 та C5.0, CART, SPRINT, ID3 та ін.

Процес індукції дерев рішень складається з двох основних етапів – вирощування дерева та обрізки дерева. Фаза росту включає рекурсивну розбивку навчальної вибірки, результатом якої є дерево рішень, у якому кожен листовий вузол асоціюється з одним класом, а при подальшій розбивці такого вузла хоча б один із нових листових вузлів опиняється за межами визначеного граничного значення. Фаза обрізки націлена на узагальнення дерева рішень шляхом створення піддерева з метою унеможливлення перенавчання. У процесі обрізки дерева рішень часто орієнтуються на вимірники точності класифікації [4, с. 262–267].

Під час побудови дерев рішень можуть використовуватися різні критерії поділу, наприклад, приріст інформації («information gain», або відстань Кульбака – Лейблера), відносна ентропія, коефіцієнт приросту інформації та ін. [4, с. 264].

Основними перевагами дерев рішень є їхня зрозумілість та краще функціонування з дискретними/категорійними даними. З іншого боку, існують проблеми застосування дерев рішень, пов'язані з використанням надвисокимірних та незбалансованих даних, статичність, обмеженість порядкової класифікації [4, с. 269].

Таблиця 2. Набори показників

Набір №	Показники
1	[C + CFI + R] / TA; E / S; II / [C + CFI + LTFI]; S / TA; QR; TL / TA
2	[C + CFI + R] / TA; E / TL; II / C+CFI + LTFI; S / TA; TL / TA; WC / S
3	[C + CFI + R] Gr; CL / TA; GP / E; CFO; NI / TA
4	[C + CFI + R] Gr; C / TA; CL / TA; CFO; RE / TA
5	[C + CFI + R] / S; C / CL; GP Gr; [C + CFI + R] / IR
6	C / CL; GP Gr; [C + CFI + R] / IR; NI / S
7	C / CL; GP Gr; [C + CFI + R] / IR; WC / S

Основні результати дослідження

Побудова моделей здійснювалася у два етапи з використанням програмного забезпечення RapidMiner 6.1. Для побудови дерев рішень використовувався алгоритм CART (Classification & Regression Trees) із вхідними наборами ознак № 1–7. Окрім цього, було побудовано дерева рішень на основі початкового переліку показників; у межах цього підходу було також застосовано алгоритм Random Forest, відповідно до якого для підвищення класифікаційних характеристик моделі будується група дерев рішень.

У процесі побудови моделей використовувалася 10-кратна крос-валідація. Оскільки дерева рішень передбачають налаштування переліку параметрів, а саме: критерію вибору атрибутів для галуження, мінімального розміру вузла

для галуження, мінімального розміру листка, мінімального приросту інформаційного критерію тощо, було використано функцію Optimize Parameters (Grid), яка дає змогу перебрати всі можливі комбінації параметрів у встановлених межах і обрати найбільш прийнятну модель залежно від обраного критерію. У таблиці 3 наведено перелік та межі параметрів, які варіювали.

Завдяки застосуванню алгоритму оптимізації було обрано найкращі з позиції певного критерію параметри дерев рішень. Результати моделювання представлено в таблиці 4.

Як видно, індикатори класифікаційних якостей побудованих моделей є доволі високими. Усі розроблені дерева рішень мають високу специфічність, чутливість та точність. Варто зауважити, що при прогнозуванні фінансової кризи на підприємстві першочергова вага має

Таблиця 3. Налаштування функції перебору параметрів

Параметр	Мінімальне значення	Максимальне значення	Кількість кроків	Шкала
Критерій галуження	–	–	4	1) відносний приріст інформації; 2) абсолютний приріст інформації; 3) індекс засмічення Джині (Gini impurity); 4) точність моделі.
Мінімальний розмір вузла	1	5	5	лінійна
Мінімальний розмір листка	1	5	5	лінійна
Мінімальний приріст критерію	0,001	1	10	квадратична

Таблиця 4. Характеристики побудованих дерев рішень

Набір змінних, №	Параметри дерева (критерій галуження; мін. розмір вузла; мін. розмір листка; мін. приріст критерію)	Специфічність, %	Чутливість, %	Точність, %
1	Абсолютний приріст інформації; 1; 1; 0,011	100,00	84,62	98,82
2	Відносний приріст інформації; 4; 1; 0,161	100,00	88,46	99,12
3	Індекс засмічення Джині; 2; 2; 0,092	100,00	80,77	98,53
4	Відносний приріст інформації; 1; 1; 0,001	100,00	92,31	99,41
5	Точність; 5; 1; 0,001	100,00	96,15	99,71
6	Точність; 1; 1; 0,001	100,00	96,15	99,71
7	Точність; 5; 1; 0,001	100,00	96,15	99,71
Повний перелік	Індекс засмічення Джині; 1; 1; 0,011	100,00	88,46	99,12
Повний перелік, Random forest	Індекс засмічення Джині; 1; 1; 0,001	100,00	96,15	99,71

Таблиця 5. Результати класифікації на основі обраних дерев рішень

Набір змінних, №	Специфічність, %	Чутливість, %	Точність, %
1	88,54	62,50	85,59
2	88,54	60,00	85,31
3	89,49	47,50	84,75
4	89,17	60,00	85,88
5	87,26	57,50	83,90
6	86,94	52,50	83,05
7	86,94	57,50	83,62
Повний перелік	87,26	50,00	83,05
Повний перелік, Random forest	96,18	37,50	89,55

надаватися здатності моделі правильно визначати потенційних банкрутів, тобто критичнішою помилкою буде назвати компанію в стані кризи фінансово здоровою, аніж навпаки. Таким чином, чутливість побудованих дерев рішень має бути переважним критерієм відбору найкращих моделей. Беручи це до уваги, можна зробити висновок, що моделі на основі наборів показників № 5–7, а також Random forest демонструють найкращі результати.

Утім, щоб пересвідчитися, що моделі мають прийнятні узагальнювальні властивості (можуть використовуватися для прогнозування), логічним кроком є перевірка їхніх класифікаційних якостей на даних наступного періоду. Результати застосування побудованих дерев рішень до даних за 2011 звітний рік наведено в таблиці 5.

Очевидно, що прогнозні якості побудованих дерев рішень значно поступаються їхнім класифікаційним характеристикам. Це може бути пов'язано з декількома факторами. Зокрема, могло відбутися перенавчання моделі, особливо враховуючи невелику кількість компаній-банкрутів у вибірці; з іншого боку, процедура крос-валідації мала обмежити цей фактор. Більш імовірною причиною відносно низьких показників точності застосування моделей на даних інших періодів є мінливість у структурі даних, а саме зміна взаємозв'язків між ознаками.

Однак серед розроблених дерев рішень найбільш прийнятні властивості має модель, побудована на основі набору змінних № 1. Дерево рішень має найвищий показник чутливості, а також прийнятні значення специфічності та загальної точності. Обране дерево рішень зображено на рисунку.

Як видно, до моделі увійшли 5 із 6 індикаторів набору № 1. Серед них такі: показник прибутковості інвестицій – $\Pi / [C + CFI + LTFI]$; частка

ліквідних активів у балансі страховика – $[C + CFI + R] / TA$; обернений показник використання власних засобів страхової компанії – E / S ; показник оборотності активів компанії – S / TA ; а також коефіцієнт швидкої ліквідності – QR . Таким чином, прогнозування фінансової кризи в українських компаніях з ризикового страхування за допомогою розробленого дерева рішень відбувається на основі аналізу рентабельності інвестиційної діяльності, ліквідності та ділової активності.

Висновки та пропозиції щодо подальших досліджень

Проведений аналіз продемонстрував, що дерева рішень можуть бути застосовані для прогнозування фінансової кризи в українських страхових компаніях. На основі декількох наборів ознак було побудовано моделі дерев прийняття рішень, які мають високі значення показників класифікаційної точності – специфічність 100 %, чутливість вище 96 % та загальну точність вище 99 %. Однак перевірка моделей на даних майбутніх періодів вказала на погіршення класифікаційних властивостей розроблених дерев рішень. Модель із найбільш прийнятними характеристиками має доволі високі значення специфічності та точності (вище 85 %), проте відносно невисоку чутливість (62,5 %), що може бути викликано властивостями звітних даних страховиків.

Вдосконалення процесу прогнозування кризових явищ в українських страхових компаніях із використанням дерев рішень може відбуватися в кількох напрямках, серед яких розширення інформаційної бази (як з позиції кількості компаній, так і з позиції кількості періодів), вдосконалення алгоритмів вибору ознак, а також формування ансамблів дерев рішень.

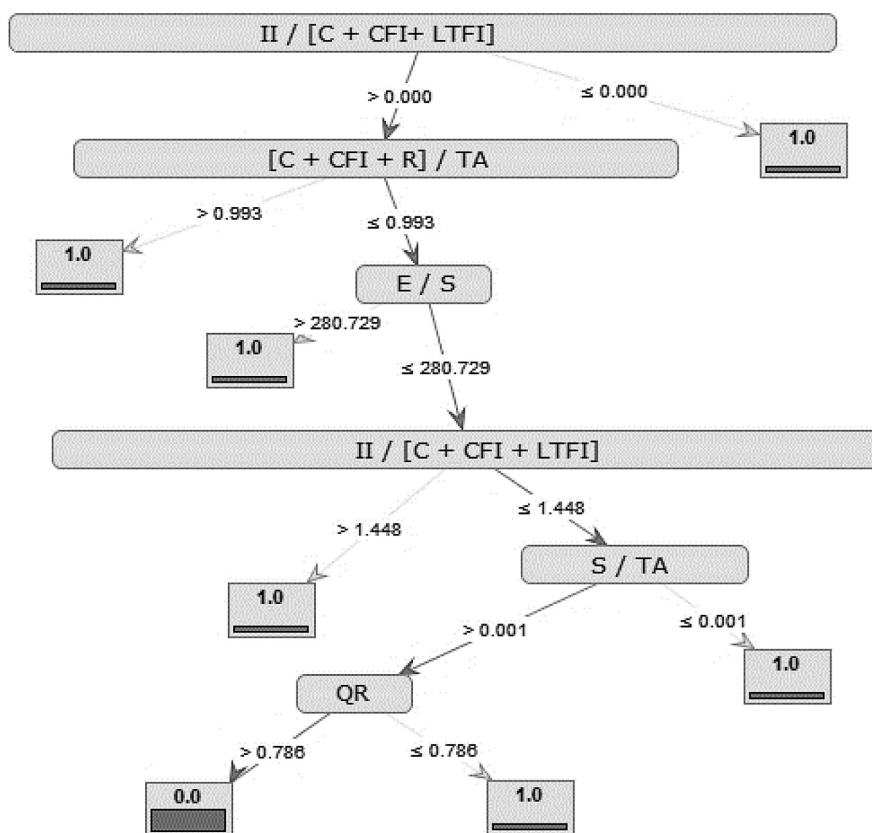


Рис. Дерево рішень з найкращими характеристиками

Список літератури

1. A comparison of neural network, statistical methods, and variable choice for life insurers' financial distress prediction / P. Brockett, L. Golden, J. Jang, Ch. Yang // The Journal of Risk and Insurance. – 2006. – № 73 (3). – P. 397–419.
2. Insolvency prediction model using artificial neural network for Malaysian general insurers / N. Chiet, S. Jaaman, N. Ismail, S. Shamsuddin // Proceedings of IEEE World congress on Nature and Biologically Inspired Computing (NaBIC), Coimbatore, India. – 2009. – P. 584–589.
3. Kleffner A. An examination of property & casualty insurer solvency in Canada / A. Kleffner, R. Lee // Journal of Insurance Issues. – 2009. – № 32 (1). – P. 52–77.
4. Kotsiantis S. Decision trees: a recent overview / S. Kotsiantis // Artificial Intelligence Review. – 2013. – № 39 (4). – P. 261–283.
5. Lee S. Analysis and prediction of insolvency in the property-liability insurance industry: a comparison of logit and hazard models / S. Lee, J. Urrutia // The Journal of Risk and Insurance. – 2006. – № 63 (1). – P. 121–130.
6. Predicting financial distress and corporate failure: a review from the state-of-the-art definitions, modeling, sampling, and featuring approaches / J. Sun, H. Li, Q.-H. Huang, K.-Y. He // Knowledge-Based Systems. – 2014. – № 57. – P. 41–56.
7. Sharpe I. Financial distress in Australian general insurers / I. Sharpe, A. Stadnik // The Journal of Risk and Insurance. – 2007. – № 74 (2). – P. 377–399.

A. Lytvyn

DEVELOPMENT OF DECISION TREES FOR FINANCIAL DISTRESS PREDICTION IN UKRAINIAN INSURANCE COMPANIES

In the article, the feasibility of applying decision trees to forecasting financial crises in Ukrainian insurance companies was analyzed. Using the formed feature sets, several decision tree models were built, among which the best performing in terms of sensitivity and accuracy model was selected based on the classification characteristics for two consecutive periods. Based on the analysis, the ways of improving the practice of applying decision trees for crisis prediction in Ukrainian insurance companies were outlined.

Keywords: decision trees, financial distress prediction, financial ratios, insurance company.

Матеріал надійшов 20.05.2015