



Юрик Я.І., канд. екон. наук

старший науковий співробітник відділу соціально-економічних проблем праці,
Інститут економіки та прогнозування НАН України

Кузьменко Г.Г., канд. екон. наук

експерт відділу моделювання ризику, АТ "Аліор банк" (Польща)

ПОБУДОВА СКОРИНГОВОЇ МОДЕЛІ ДЛЯ ОЦІНКИ РИЗИКОВИХ ПОДІЙ НА РИНКУ ПРАЦІ

Уперше досліджено можливості побудови скорингової моделі для розв'язання завдань прогнозування, класифікації та діагностики ризику на ринку праці, зокрема безробіття.

Розроблення скорингової моделі здійснювалося за допомогою методу бінарної логістичної регресії з використанням програм R, SPSS та MS Excel.

Результат, який дозволила отримати технологія скорингу, – ранжування зайнятих за "потенціалом" втрати ними роботи та виокремлення у структурі робочої сили контингенту, для якого цей ризик більший.

Використання такого результату створить підґрунтя для вироблення рішень з подолання проблем безробіття не інтуїтивно, а на основі формалізованих критеріїв, безпосередньо пов'язаних з імовірністю настання цього ризику.

К л ю ч о в і с л о в а : скорингова модель, ризик, ринок праці, безробіття.

J E L : C25, E24, E27

Ризики, на які неминуче наражаються суб'єкти на ринку праці, змушують шукати різні способи попередження та мінімізації їхніх наслідків. Водночас завдання управляти ризиком актуалізує аналіз можливостей його передбачення.

Сьогодні одним із найуспішніших прикладів використання математичних і статистичних методів діагностики і прогнозування ризику є скоринг. Застосовується він зараз переважним чином у банківській діяльності. Та поряд із уже традиційним його використанням¹ є актуальним перенесення відповідного досвіду на вирішення завдань і в інших сферах. Зокрема, ця робота досліджує можливості побудови скорингової моделі для оцінки ризикових подій на ринку праці.

Так, з урахуванням наявної доступної інформації **за мету було поставлено** виявити, яким чином соціально-демографічні та професійно-кваліфікаційні характеристики зайнятих в Україні впливають на ймовірність їхнього переходу у статус безробітних (аплікаційний скоринг). Результати такого моделювання даватимуть базис для розпізнавання образів та автоматичної класифікації працівників за "потенціалом" втрати ними роботи, що в свою чергу сприятиме спрямуванню вироблення превентивних заходів із обмеження проблем безробіття в Україні.

Аналіз публікацій дає підстави вважати, що робіт, присвячених кредитному скорингу – як теоретико-методичного, так і практичного спрямування – сьогодні чимало й їхня кількість продовжує зростати [2–18 та ін.]. Оприлюднені результати

¹Так, за допомогою скорингу фінансовій установі забезпечується оцінка кредитоспроможності позичальників (англ. *application scoring* – аплікаційний скоринг, або скоринг заявника); оцінка ймовірності повернення виданих кредитів (англ. *behavioral scoring* – поведінковий скоринг); оцінка можливості повного або часткового повернення позики в разі порушення термінів погашення заборгованості (англ. *collection scoring*); оцінка ймовірності того, що новий клієнт не є шахраєм (англ. *fraud scoring*); оцінка реакції споживача на спрямування йому пропозиції (англ. *response scoring*); оцінка ймовірності подальшого використання банківського продукту або перехід до іншого постачальника (англ. *attrition scoring*) [1].

щодо можливостей застосування скорингових моделей у відмінних від банківської сферах переважним чином зосереджені на клінічній практиці та маркетингу [19–22 та ін.]. Стосовно ж ринку праці релевантні дослідження відсутні. Тому вперше апробовано створення скорингової системи оцінки ризику безробіття серед зайнятого населення України, а результати представлено у цій статті.

Основна частина. Сучасний арсенал методів побудови скорингових моделей забезпечується використанням різноманітних інструментів предиктивного аналізу (англ. *prediction* – прогноз, передбачення), що належать до широкого класу технологій поглибленого аналізу даних (*data science*).

До основних інструментів предиктивного аналізу відносять: статистичні методи (лінійну множинну та логістичну регресію); дерево класифікації або рекурсивно-partiційний алгоритм; нейромережі [1].

Нейромережі зазвичай критикують за непрозорість системи прийняття рішення, на відміну від дерев класифікації, які демонструють чітку логіку розрахунків, проте не дають можливості використовувати вхідні кількісні змінні. Лінійну регресію, навпаки, краще застосовувати у випадку кількісних змінних. Логістична ж регресія і нейронні мережі ефективні за будь-яких даних – як якісних, так і кількісних.

В основі розробленої нами скорингової моделі лежить метод бінарної логістичної регресії, який дозволяє виявляти залежність дихотомічних змінних від цілого ряду незалежних факторів. Серед останніх, відповідно до мети дослідження та теоретичних гіпотез, обрано такі характеристики працівника, як вік, місце проживання, гендер, сімейний стан, рівень освіти, отриманий фах, професія, сфера діяльності. Залежна змінна набуває лише двох значень (у нашому випадку – безробітний/зайнятий) та має біноміальний розподіл (табл. 1).

Для побудови будь-якої скорингової моделі необхідно мати інформаційну базу, яка буде, по-перше, достатньою за обсягом та якістю, а, по-друге, матиме певну історичну давність. З метою дослідження цим критеріям відповідали первинні дані щомісячних вибіркового обстеження населення (домогосподарств) з питань економічної активності за 2010–2013 рр. А вже 2014–2015 рр. – часовий інтервал, який доцільно використати саме для моніторингу – процедури, спрямованої на діагностику адекватності функціонування розробленої моделі.

Зауважимо, що на момент підготовки публікації результати вибіркового обстеження населення з питань економічної активності за 2015 р. були відсутні, крім того, метою представленої роботи була насамперед апробація побудови скорингу в нетрадиційному для цього середовищі. А, отже, моніторинг розглянемо у наших наступних наукових дослідженнях.

Ще до того, як буде розроблена скорингова модель, необхідно дати точне означення "поганого" (*bad*) випадку, а всю інформаційну базу:

по-перше, розділити на дві групи: основну вибірку – спостереження, за якими власне й буде створюватися модель, та тестову – дані, які не братимуть участі в моделюванні, а використовуватимуться для первинної валідації – перевірки якості прогнозних можливостей моделі до початку її використання. Основна та тестова вибірки формуються на основі механізму випадкового відбору і, як правило, у співвідношенні 70–80% і 30–20% відповідно до початкового об'єму загальної сукупності даних;

по-друге, перетворити у придатну для подальшого аналізу форму. Відомо два основні підходи, які використовуються при роботі як із кількісними, так і з якісними змінними. Перший – це перетворення кожного вихідного значення ознаки в окрему



Методи і моделі прогнозування

бінарну змінну. Але такий підхід, хоч методологічно і простий, проте незручний тим, що зумовлює введення великого числа змінних. За другим підходом, який і було використано, здійснюється перетворення кожного вихідного значення ознаки у величину, що дорівнює логарифму відношення відповідних відсотків "хороших" випадків до "поганих". У результаті кожне з вихідних значень ознаки набуває числового значення, що відповідає рівню його "ризиковості".

Таблиця 1

Набір можливих змінних скорингової моделі для прогнозування ризику безробіття серед зайнятого населення України

Характеристики працівника	Вихідні значення ознаки (характеристики працівника)
Статус економічної активності	1 – безробітний, або "bad" 0 – зайнятий, або "good"
Вік	Кількісна змінна: від 15 до 59 років ² .
Місце проживання	Якісна змінна: 1 – міські поселення; 2 – сільська місцевість
Стать	Якісна змінна: 1 – жінка; 2 – чоловік
Сімейний стан	Якісна змінна: 1 – перебуває у шлюбі; 2 – не перебуває у шлюбі; 3 – розлучений(а); 4 – удівець (вдова); 5 – діти до 18 років, що не перебувають у шлюбі.
Рівень освіти	Якісна змінна: 1 – повна вища; 2 – базова вища; 3 – неповна вища; 4 – повна загальна середня; 5 – базова загальна середня; 6 – початкова загальна; 7 – немає початкової.
Професія, спеціальність згідно з посвідченням (дипломом)	Якісна змінна: 1 – законодавці, вищі державні службовці, керівники, менеджери (управителі); 2 – професіонали; 3 – фахівці; 4 – технічні службовці; 5 – працівники сфери торгівлі та послуг; 6 – кваліфіковані робітники сільського та лісового господарств, риборозведення та рибальства; 7 – кваліфіковані робітники з інструментом; 8 – робітники з обслуговування, експлуатації та контролювання за роботою технологічного устаткування, складання устаткування та машин; 9 – найпростіші професії; 10 – закінчили курси; 11 – не мають професії.
Професія за основною роботою	Якісна змінна: 1 – законодавці, вищі державні службовці, керівники, менеджери (управителі); 2 – професіонали; 3 – фахівці; 4 – технічні службовці; 5 – працівники сфери торгівлі та послуг; 6 – кваліфіковані робітники сільського та лісового господарств, риборозведення та рибальства; 7 – кваліфіковані робітники з інструментом; 8 – робітники з обслуговування, експлуатації та контролювання за роботою технологічного устаткування, складання устаткування та машин; 9 – найпростіші професії.
Вид діяльності за основною роботою	Якісна змінна: 1 – сільське господарство, мисливство, лісове господарство; 2 – рибальство, рибицтво; 3 – добувна промисловість; 4 – виробництво та розподілення електроенергії, газу та води; 6 – будівництво; 7 – оптова й роздрібна торгівля; 8 – діяльність готелів та ресторанів; 9 – діяльність транспорту та зв'язку; 10 – фінансова діяльність; 11 – операції з нерухомим майном; 12 – державне управління; 13 – освіта; 14 – охорона здоров'я; 15 – інші види економічної діяльності (в т.ч. надання комунальних та індивідуальних послуг; діяльність домашніх господарств; діяльність екстериторіальних організацій; робота за кордоном; діяльність домашніх господарств як виробників товарів та послуг для власного споживання).

Джерело: складено авторами за даними вибіркового обстеження населення (домогосподарств) з питань економічної активності.

Параметри підготовленої з метою дослідження інформаційної бази узагальнено в табл. 2.

Як зазначалося вище, наступним етапом – перед безпосереднім моделюванням – є аналіз скорингових змінних, а саме групування, перевірка їхньої прогностичної сили, діагностика мультиколінеарності.

² За правилами скорингового моделювання до відібраної сукупності даних не повинні входити нестандартні випадки, через що інформацію, яка стосувалася зайнятих у віці 60–70 р., вилучено.

Таблиця 2

Основні параметри інформаційної бази для прогнозування ризику безробіття серед зайнятого населення України

Тривалість періоду спостережень для побудови моделі	2010–2013 рр.	
Період для моніторингу функціонування моделі*	2014–2015 рр.	
Визначення "поганого" (<i>bad</i>) випадку для вибірки	Перехід зайнятого у статус безробітного хоча б раз упродовж наступних 14 ³ місяців	
Опис інформаційної бази**		
Характеристики	Основна вибірка (80%)	Тестова вибірка (20%)
Кількість спостережень	19514	4866
Кількість <i>bad</i>	815	200
<i>Bad rate</i>	4,18%	4,11%

* Як зазначалось вище, через об'єктивні обставини моніторинг буде розглянуто у наступних наукових дослідженнях.

** Підготовку вибірки здійснено за допомогою **R** – мови програмування для статистичної обробки даних та роботи з графікою.

Джерело: авторські розрахунки.

Групування всередині кожної змінної (особливо це стосується тих, що з великою кількістю вихідних значень) дозволяє легше зрозуміти існуючі залежності в моделі, а кінцеві скорингові бали будуть більш транспарентними та стійкими до незначних змін в інформаційній базі. Крім того, групування вирішує проблему екстремальних величин і рідкісних значень та надає більшої статистичної значимості скоринговими змінним.

Для оцінки прогностичної сили скорингових змінних розраховують показник інформаційного значення (англ. *information value*, *IV*). Чим вище інформаційне значення змінної, тим більшу вагу вона має з точки зору корисності при побудові моделі.

За розрахунками виявлено (табл. 3), що попри різну предиктивну силу⁴, всі фактори можуть бути відібраними для скорингу (*IV* у кожному випадку більше 0,02), а найбільш значимими у передбаченні ймовірності настання "поганого" випадку є такі характеристики працівника, як вік, професія (фактична) та вид діяльності за основною роботою.

Таблиця 3

Оцінка прогностичної сили змінних у ймовірнісному моделюванні ризику безробіття серед зайнятого населення України

№ п/п	Змінна	<i>IV</i>
1.	Вік працівника	0,156
2.	Професія за основною роботою	0,147
3.	Вид діяльності за основною роботою	0,129
4.	Місце проживання	0,094
5.	Сімейний стан	0,087
6.	Рівень освіти	0,083
7.	Професія, спеціальність згідно з посвідченням (дипломом)	0,083
8.	Гендер	0,041

Джерело: авторські розрахунки.

³ 14 місяців – такий термін зумовлений особливостями проведення вибірових обстежень населення (домогосподарств) з питань економічної активності: щомісячний обсяг вибіркової сукупності домогосподарств формується з урахуванням схеми ротації, за якою кожне відібране домогосподарство опитують 6 разів: 3 місяці поспіль опитування – перерва у 9 місяців – опитування 3 місяці поспіль.

⁴У професійному середовищі існує така градація змінних за рівнем *IV*: менше 0,02 – незалежна змінна не корисна для прогнозування; 0,02–0,1 – низька прогностична сила; 0,1–0,3 – середня прогностична сила; 0,3 і більше – висока прогностична сила.



Методи і моделі прогнозування

Одна з найважливіших умов методу логістичної регресії – відсутність мультиколінеарності. За її наявності необхідно знайти оптимальні варіанти виключення тісно корельованих змінних для побудови моделі.

Аналіз на основі матриці парних кореляцій продемонстрував існування тісного зв'язку між деякими нашими факторами (табл. 4).

Таблиця 4

Матриця парних кореляцій між змінними скорингової моделі для прогнозування ризику безробіття серед зайнятого населення України

Змінна	Місце проживання	Професія, спеціальність згідно з посвідченням (дипломом)	Сімейний стан	Гендер	Рівень освіти	Вік працівника	Професія за основною роботою	Вид діяльності за основною роботою
Місце проживання	1	0,237	0,021	0,012	0,296	0,002	0,290	0,271
Професія, спеціальність згідно з посвідченням (дипломом)	0,237	1	-0,019	0,202	0,932	-0,014	0,688	0,235
Сімейний стан	0,021	-0,019	1	0,181	-0,008	0,418	0,010	0,073
Гендер	0,012	0,202	0,181	1	0,122	0,029	0,253	0,215
Рівень освіти	0,296	0,932	-0,008	0,122	1	-0,001	0,692	0,253
Вік працівника	0,002	-0,014	0,418	0,029	-0,001	1	-0,023	0,004
Професія за основною роботою	0,290	0,688	0,010	0,253	0,692	-0,023	1	0,428
Вид діяльності за основною роботою	0,271	0,235	0,073	0,215	0,253	0,004	0,428	1

Джерело: авторські розрахунки.

Отже, до моделі насамперед увійдуть такі змінні, як вік працівника, його місце проживання, сімейний стан, гендер, професія згідно з дипломом та вид діяльності за основною роботою. А з метою усунення мультиколінеарності та максимального збереження інформаційного навантаження, яке несуть такі фактори, як рівень освіти та фактична професія, і, тим самим, – збільшення прогностичної сили майбутньої моделі, було введено нову змінну "чи виконував працівник роботу за отриманим фахом". Відповідь на це питання (так/ні) у поєднанні з даними щодо професії, отриманої згідно з дипломом, дозволяє сформулювати певне уявлення як про освітній рівень формальної підготовки працівника, так, власне, і про фактично виконувану ним роботу.

Параметри рівняння логістичної регресії для прогнозування ризику безробіття серед зайнятого населення України представлено в табл. 5.

Разом із інформацією щодо стандартного відхилення за кожною з незалежних змінних коефіцієнти логістичної регресії дозволяють оцінити та порівняти відносний вклад кожного з предикторів у модель (табл. 6).

Отже, найбільшу вагу в імовірності настання "поганого" випадку має фактор віку працівника (26%), найменшу – гендер (7,8%).

Важливою характеристикою будь-якої статистичної моделі є її достовірність, яка у випадку логістичної регресії характеризується здатністю останньої відрізнити "хороші" випадки від "поганих". Для оцінки якості класифікаційних можливостей моделі використовують насамперед індекс Джині, тест Колмогорова – Смірнова та площу під ROC-кривою.

Таблиця 5

Параметри рівняння логістичної регресії для прогнозування ризику безробіття серед зайнятого населення України*

Змінна	Coefficient (B)	Wald statistic	Significance level of the Wald statistic**	Estimated odds ratio (exp(B))
Вік працівника	-0,785	58,772	0,0000	0,456
Вид діяльності за основною роботою	-0,543	21,618	0,0000	0,581
Місце проживання	-0,577	20,247	0,0000	0,561
Професія, спеціальність згідно із дипломом	-0,614	18,254	0,0000	0,541
Відповідність фактичної зайнятості отриманому фаху	-0,572	10,274	0,0010	0,565
Сімейний стан	-0,348	6,189	0,0130	0,706
Гендер	-0,459	5,839	0,0160	0,632
Константа	-3,129	7010,158	0,0000	0,044

*Статистичну оцінку параметрів рівняння логістичної регресії проведено за допомогою статистичного пакету SPSS 20.

** Якщо рівень значимості статистики Wald менший за 0,05, то змінна корисна для моделі.

Джерело: авторські розрахунки.

Таблиця 6

Ефективні ваги незалежних змінних у моделі прогнозування ризику безробіття серед зайнятого населення України

Змінна	B	Стд. відхил.	Ефективні ваги незалежних змінних, %
Вік працівника	-0,785	0,3987	26,00
Вид діяльності за основною роботою	-0,543	0,3853	17,38
Місце проживання	-0,614	0,2990	15,25
Професія, спеціальність згідно з посвідченням (дипломом)	-0,577	0,3051	14,63
Відповідність фактичної зайнятості отриманому фаху	-0,572	0,2299	10,93
Сімейний стан	-0,348	0,2770	8,01
Гендер	-0,459	0,2046	7,80

Джерело: авторські розрахунки.

З огляду на те, що побудована логістична регресія є апікаційною моделлю з відносно невеликою кількістю входних змінних, отримані значення показників (табл. 7) відповідають хорошій якості бінарного класифікатора.

Таблиця 7

Показники оцінювання якості моделі для прогнозування ризику безробіття серед зайнятого населення України

Показник	Основна (80%)	Тестова (20%)
Коефіцієнт Джині (Gini) ⁵	34,6%	33,6%
Тест Колмогорова – Смірнова ⁶	26,1%	23,0%
Площа під ROC-кривою (AUC ⁷)	0,673	0,668

Джерело: авторські розрахунки.

Заключним етапом розробки скорингової моделі є масштабування (калібрація) – прийом переведення первинних скорингових балів у шкалу, яка зручніша для користування, зокрема, з діапазоном від 0 до 1000. Результати цього процесу трансформуються у скорингову карту (табл. 8).

⁵ Коефіцієнт Джині переводить значення площі під ROC-кривою в діапазон від 0 до 1 (або 0–100%).

⁶ Діапазон значень статистики Колмогорова – Смірнова перебуває від 0 до 100%.

⁷ Розрахункове значення показника площі під ROC-кривою – AUC – може перебувати у межах від 0,5 до 1.

Скорингова карта прогнозування ризику безробіття
серед зайнятого населення України

Змінна	Значення	SCORE
Вік працівника	15–22	44
	23–25	68
	26–47	83
	48–51	98
	52–55	110
	56–59	130
Вид діяльності за основною роботою	Сільське, лісове та рибне господарства; Будівництво	74
	Переробна промисловість; Фінансова діяльність; Операції з нерухомістю та надання послуг підприємцям; Інші види економічної діяльності	81
	Державне управління	85
	Торгівля, діяльність готелів та ресторанів; Транспорт та зв'язок	92
	Добувна промисловість; Виробництво та розподілення електроенергії, газу та води; Освіта	102
	Охорона здоров'я та надання соціальної допомоги	114
Місце проживання	Сільська місцевість	75
	Міські поселення	95
Професія, спеціальність згідно із посвідченням (дипломом)	Кваліфіковані робітники сільського та лісового господарств, риборозведення та рибальства; Кваліфіковані робітники з інструментом; Робітники з обслуговування, експлуатації та контролювання за роботою технологічного устаткування, складання устаткування та машин	73
	Технічні службовці; Працівники сфери торгівлі та послуг; Найпростіші професії; Закінчили курси; Без професії	81
	Фахівці	89
	Законодавці, вищі державні службовці, керівники, менеджери (управителі); Професіонали	103
Виконував роботу за отриманим фахом	Ні	80
	Так	96
Сімейний стан	Неодружений	74
	Одружений	87
	Розлучений(а); удівець (удова)	93
Гендер	Чоловік	80
	Жінка	91

Джерело: авторські розрахунки.

Підсумком калібрації моделі є присвоєння кожному працівнику рейтингового класу за рівнем ризику настання "поганого" випадку. З цією метою було підібрано відповідну шкалу, яка найкращим чином відображає побудовану модель (табл. 9).

Якість калібрації можна перевірити за допомогою тестів, завданням яких є встановити, наскільки істинні значення досліджуваного показника відповідають прогнозованим.

Одним із таких засобів діагностики є тест з використанням стандартного нормального розподілу. Перевірка здійснюється для кожного рейтингового класу та в цілому й може бути односторонньою або ж двосторонньою. З точки зору управління, актуальним перш за все є ризик недопрогнозування ймовірності втратити роботу, тому в табл. 10 представлено результати виявлення саме цих значимих відхилень у калібруванні побудованої моделі.

Альтернативним тестом на відповідність розрахованої ймовірності та фактичного безробіття є біноміальний (табл. 11).

Отже, за результатами перевірки (табл. 10 та 11) можна зробити висновок про хорошу якість калібрації моделі, а відхилення спостерігаються лише в одному з семи класів.

Таблиця 9

Шкала для визначення рейтингового класу працівника за рівнем ризику настання "поганого" випадку

Сумарний бал	Імовірність того, що зайнятий упродовж наступних 14-ти місяців хоча б раз буде безробітним	Рейтинговий клас
(0; 530]	11,67% і більше	A7
(530; 560]	7,28–11,67%	A6
(560; 590]	4,46–7,28%	A5
(590; 620]	2,70–4,46%	A4
(620; 650]	1,62–2,70%	A3
(650; 680]	0,97–1,62%	A2
(680; 1000]	0–0,97%	A1

Джерело: авторські розрахунки.

Таблиця 10

Перевірка якості калібрації моделі для прогнозування ризику безробіття серед зайнятого населення України

Рейтинговий клас	Bad rate _{факт}	Bad rate _{прогноз}	BR _ф – BR _п	Область допустимих значень для рівня довіри:			
				90%	95%	99%	99,9%
A7	13,64%	14,60%	-0,97%	2,06%	2,64%	3,73%	4,96%
A6	6,89%	8,37%	-1,48%	0,85%	1,09%	1,54%	2,05%
A5	6,53%	5,54%	0,99%	0,43%	0,55%	0,77%	1,03%
A4	3,41%	3,45%	-0,05%	0,30%	0,38%	0,54%	0,71%
A3	2,00%	2,19%	-0,19%	0,28%	0,36%	0,51%	0,68%
A2	1,17%	1,36%	-0,19%	0,37%	0,47%	0,67%	0,89%
A1	0,00%	0,81%	-0,81%	0,72%	0,93%	1,31%	1,74%
Всі класи	4,18%	4,18%	0,00%	0,18%	0,24%	0,33%	0,44%

Джерело: авторські розрахунки.

Таблиця 11

Виявлення значимих відхилень в калібруванні моделі для прогнозування ризику безробіття серед зайнятого населення України (біноміальний тест)

Рейтинговий клас	Кумулят. біном. розподіл	Результат тесту з рівнем довіри:			
		90%	95%	99%	99,9%
A7	29,94%				
A6	1,28%				
A5	99,84%	Недооцінено	Недооцінено	Недооцінено	
A4	43,90%				
A3	20,99%				
A2	29,80%				
A1	12,88%				
Усі класи	49,63%				

Джерело: авторські розрахунки.

Ранжування зайнятих за рівнем ризику настання "поганого" випадку відображено на рис. 1, а узагальнену інтерпретацію скоркарти (табл. 8) подано в таблиці 12.

Якщо детальніше, то ситуація щодо впливу соціально-демографічних і професійно-кваліфікаційних характеристик зайнятих України на ймовірність їхнього переходу у статус безробітних (табл. 8), а також її пояснення виглядають таким чином.

Зайняті у віці 15–22 роки – здебільшого працюючі студенти (учні) – частіше з-поміж інших вікових категорій звільнюються як добровільно у пошуках тимчасового підробітку або ж "свого" робочого місця, так і з ініціативи роботодавця, якому звільнити "новеньких" значно простіше, аніж тих, хто тривалий час працював на благо компанії.

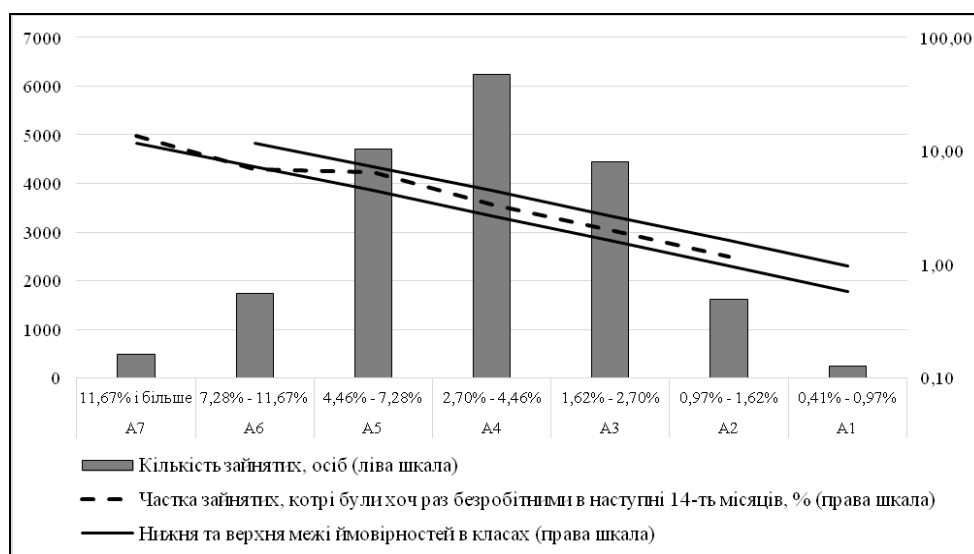


Рис. 1. Рейтингова класифікація зайнятих залежно від рівня ризику їхнього переходу у статус безробітних

Джерело: авторська розробка.

Таблиця 12

Інтерпретація результатів прогнозування ризику безробіття серед зайнятого населення України

Характеристика	Потенційні безробітні*
Вік	Чим молодший, тим більша ймовірність
Вид діяльності за основною роботою	Для зайнятих у сільському господарстві або на будівництві ризик вищий
Місце проживання	У сільській місцевості ризик більший
Професія, спеціальність згідно із посвідченням (дипломом)	Кваліфіковані робітники сільського та лісового господарств, риборозведення та рибальства; Кваліфіковані робітники з інструментом; Робітники з обслуговування, експлуатації та контролювання за роботою технологічного устаткування, складання устаткування та машин – професії з вищим ризиком
Професія за основою роботою	У зайнятих не за отриманим фахом ризик вищий
Сімейний стан	У неодружених ризик більший
Стать	У працівників чоловічої статі ризик вищий

* Нагадаємо, що за категорією "потенційні" безробітні маємо на увазі тих зайнятих, які упродовж наступних 14 місяців хоча б раз можуть стати безробітними.

Джерело: авторська розробка.

Кваліфіковані робітники сільського та лісового господарств; кваліфіковані робітники з інструментом; робітники з обслуговування, експлуатації та контролювання за роботою технологічного устаткування, а також сільське господарство та будівництво – це отриманий фах і види діяльності, для яких більше, ніж для інших, є характерним або низький попит на ринку праці, або непостійний характер зайнятості, а, отже, більшою є ймовірність безробіття.

Чоловіки менш консервативні щодо добровільної зміни роботи, їхня зайнятість частіше, ніж у жінок, має непостійний характер (випадкова, тимчасова, сезонна). Крім того, звільнення чоловіків простіше і з ініціативи роботодавців, котрим не потрібно враховувати законодавчих обмежень, які мають місце при звільненні жінок.

Неодружені перебувають у пошуку – і не лише своєї другої половинки, а й кращої роботи, тому частіше з-поміж інших схильні до звільнення за власним бажанням, а, оскільки це працівники переважно без дітей, то й звільнення з ініціативи роботодавця не має додаткових обмежень.

Звідси, найбільша ймовірність (18,2%) хоча б раз на 14 місяців стати безробітним – у працівників чоловічої статі, віком 15–22 роки, не одружених, які проживають у сільській місцевості, з професією (спеціальністю) – згідно з дипломом (посвідченням) – кваліфікованих робітників сільського та лісового господарств; кваліфікованих робітників з інструментом; робітників з обслуговування, експлуатації та контролю за роботою технологічного устаткування, при цьому зайнятих не за фахом, а, точніше, тих, які виконують найпростіші роботи у таких видах економічної діяльності, як сільське господарство або будівництво.

Висновки

Таким чином, уперше апробовано створення скорингової моделі для розв'язання завдань прогнозування, класифікації та діагностики ризикових подій на ринку праці. Зокрема, побудовано модель, яка, враховуючи спільний вплив соціально-демографічних і професійно-кваліфікаційних характеристик зайнятих, розраховує бали, на підставі яких ранжує працівників за рівнем ризику (ймовірністю) втрати ними роботи (безробіття).

Результати тестування якості розробленої моделі, а також емпіричні оцінки, що підтверджують теоретичні припущення, фактично доводять можливість і перспективність застосування скорингу для діагностики та передбачення ризикових подій на ринку праці.

На базі розробленої моделі не тільки структурується процес підготовки можливих варіантів рішень з управління ризиком, а й проводиться попередня оцінка значимості досліджуваних характеристик зайнятого, пов'язаних із ймовірністю настання "поганої" події.

Що ж до сфери застосування скорингових моделей з оцінки ризикових подій на ринку праці (безробіття, в т.ч. довготривалого, нестандартної, в т.ч. неформальної зайнятості та ін.), то актуальним такий інструмент може бути як для державних інститутів, скажімо служби зайнятості, так і для роботодавців, тобто суб'єктів ринку праці, у яких щодня збираються і нагромаджуються великі обсяги даних, з котрих можна "витягувати" знання, та приймати рішення не інтуїтивно, а на їхній основі.

Перспективність скорингового моделювання для потреб роботодавців пов'язана й з особливістю інформаційної бази останніх, яка робить доступним виконання не лише аплікаційного скорингу, а й поведінкового, котрий враховуватиме у тому числі й мотиваційні фактори, що визначають трудову поведінку працівників та пов'язані з ймовірністю настання відповідної досліджуваної події.

Список використаних джерел

1. Юринець Р.В. Економетрична модель оцінювання кредитного позичальника відповідно до експертної оцінки / Р.В. Юринець // Науковий вісник НЛТУ України : зб. наук.-техн. праць. – Вип. 19.5. – Львів : РВВ НЛТУ України, 2009. – С. 254–258.
2. Камінський А.Б. Скорингові технології в кредитному ризик-менеджменті / А.Б. Камінський, К.К. Писанець // Бізнес Інформ. – 2012. – № 4. – С. 197–201.
3. Ковалев М. Методика построения банковской скоринговой модели для оценки кредитоспособности физических лиц / М. Ковалев, В. Корженевская // Банки Казахстана. – № 1. – 2008. – С. 43–48.
4. Крістіюгло Г.М. Використання скорингових моделей в умовах невизначеності та ризику споживчого кредитування / Г.М. Крістіюгло // Формування ринкових відносин в Україні. – 2007. – № 7 (74). – С. 86–90.
5. Сонникова В.И. Модель кредитного скоринга как алгоритм типологии нечетких совокупностей / В.И. Сонникова, К.О. Кулиджоглян // Вестник НГУЭУ. – 2013. – № 3. – С. 107–117.
6. Уланов С.В. Оценка качества и сравнение скоринговых карт / С.В. Уланов // Экономические науки. – 2009. – № 9 (58). – С. 330–335.
7. Abdou H.A., Pointon J. Credit scoring, statistical techniques and evaluation criteria: A review of the literature // Intelligent systems in accounting, finance and management. – 2011. – № 18 (2–3). – P. 59–88.
8. Anderson R. The credit scoring toolkit: theory and practice for retail credit risk management and decision automation. – New York : Oxford University press, 2007. – 790 p.



9. Hand D.J., Henley W.E. Statistical classification methods in consumer credit scoring: a review // Journal of the Royal Statistical Society, Series A. – 1997. – V. 160. – P. 523–541.
10. Liu Y. New issues in credit scoring application // Research paper, Institute of Information Systems, University of Goettingen. – 2001. – № 16. – 35 p.
11. Liu Y. A framework of data mining application process for credit scoring // Research paper, Institute of Information Systems, University of Goettingen. – 2002. – № 1. – 45 p.
12. Liu Y. The evaluation of classification models for credit scoring // Research paper, Institute of Information Systems, University of Goettingen. – 2002. – № 2. – 59 p.
13. Lewis E.M. An introduction to credit scoring. – San Rafael : The Athena Press, 1992. – 172 p.
14. Mays E. (Ed.). Handbook of credit scoring. – Chicago : Glenlake Publishing Company Ltd/Fitzroy Dearborn Publishers, 2001. – 382 p.
15. Siddiqi N. Credit risk scorecards: developing and implementing intelligent credit scoring. – New Jersey : John Wiley and Sons, 2006. – 208 p.
16. Sohn S.Y., Kim D.H., Yoon J.H. Technology credit scoring model with fuzzy logistic regression // Applied Soft Computing. – 2016. – Vol. 43. – P. 150–158.
17. Tomas L.C. A survey of credit and behavioral scoring: forecasting financial risk of lending to consumers // International Journal of Forecasting. – 2000. – Vol. 16. – P. 149–172.
18. Thomas L.C., Edelman D.B., Crook J.N. Credit scoring and its applications: SIAM monographs on mathematical modeling and computation. – SIAM : Philadelphia, USA. – 2002. – 248 p.
19. Kebebew E, Hwang J., Reiff E., Duh QY, Clark O.H. Predictors of single-gland vs multigland parathyroid disease in primary hyperparathyroidism: a simple and accurate scoring model // Archives of Surgery. – 2006. – Vol. 141 (8). – P. 777–782.
20. Мильчаков К.С. Скоринговые карты в медицине: обзор и анализ публикаций / К.С. Мильчаков, М.П. Шебалков // Врач и информационные технологии. – 2015. – № 1. – С. 71–79.
21. Malthouse E.C. Ridge regression and direct marketing scoring models // Journal of Interactive Marketing. – 1999. – Vol. 13 (4). – P. 10–23.
22. Malthouse E.C. Assessing the performance of direct marketing scoring models // Journal of Interactive Marketing. – 2001. – Vol. 15 (1). – P. 49–62.

Надійшла до редакції 12.08.2016 р.

Юрик Я.И., канд. екон. наук

старший научный сотрудник отдела социально-экономических проблем труда,

Институт экономики и прогнозирования НАН Украины

Кузьменко Г. Г., канд. екон. наук

эксперт отдела моделирования рисков, АО "Алиор банк" (Польша)

ПОСТРОЕНИЕ СКОРИНГОВОЙ МОДЕЛИ ДЛЯ ОЦЕНКИ РИСКОВЫХ СОБЫТИЙ НА РЫНКЕ ТРУДА

Впервые исследованы возможности построения скоринговой модели для решения задач прогнозирования, классификации и диагностики рисков на рынке труда, в частности безработицы.

Разработка скоринговой модели осуществлялась с помощью метода бинарной логистической регрессии с использованием программ R, SPSS и MS Excel.

Результат, который позволила получить технология скоринга, – ранжирование занятых по "потенциалу" потери ими работы и выделение в структуре рабочей силы контингента, для которого этот риск больший.

Использование подобного результата создаст основу для выработки решений относительно преодоления проблем безработицы не интуитивно, а на основе формализованных критериев, непосредственно связанных с вероятностью наступления этого риска.

Ключевые слова: скоринговая модель, риск, рынок труда, безработица.

Y. Yuryk,

PhD in Economics, Senior Researcher, Institute for Economic and Forecasting, NAS of Ukraine

G. Kuzmenko,

PhD in Economics, Expert of Retail Risk Modelling, Alior Bank S.A., Poland

CREATING A SCORING MODEL TO ASSESS RISK EVENTS ON THE LABOR MARKET

The study's main goal is testing the creation of a scoring model to meet the challenges of forecasting, classification and diagnosis of risk events on the labor market. Particularly, the given model includes joint influence of socio-demographic and skill based characteristics of the employees, provides a point based ranking of the workers by the risk (probability) of the loss of job (unemployment).

The given model not only allows structuring the process of preparing possible solutions for the risk management, but also provides a basis for the preliminary assessment of significance of the employees' characteristics, supposedly related to the probability of their transition to the unemployed.

The use of scoring models as a tool of assessing the risk events in the labor market may be useful both for public institutions, for example employment services, and for the employers.

Keywords: scoring model, risk, labor market, unemployment.

References

1. Yurynets, R.V. (2009). Ekonometrychna model' otsiniuvannia kredytnoho pozychal'nyka vidpovidno do ekspertnoi otsinky [Econometric model of evaluation of credit borrower accordingly to expert estimation]. *Naukovyj visnyk NLTU Ukrainy – Scientific Bulletin of Ukrainian National Forestry University*, 19.5, 254–258 [in Ukrainian].
2. Kaminskyi, A.B., Pysanets, K.K. (2012). Ckorynhovi tekhnolohii v kredytnomu ryzyk-menedzhmenti [Scoring technologies in credit risk-management]. *Biznes Inform – Business Inform*, 4, 197–201 [in Ukrainian].
3. Kovalev, M., Korzhenevskaya, V. (2008). Metodika postroeniya bankovskoy skoringovoy modeli dlja ocenki kreditosposobnosti fizicheskikh lic [Method of construction a bank scoring model to assess the creditworthiness of individuals]. *Banki Kazahstana – Banks of Kazakhstan*, 1, 43–48 [in Russian].
4. Kristiohlo, G. M. (2007). Vykorystannia skorynhovykh modelej v umovakh nevyznachenosti ta ryzyku spozhychoho kredyтуvannia [Using scoring models under uncertainty and risk of consumer credit]. *Formuvannia rynkovykh vidnosyn v Ukraini – Formation of Market Relations in Ukraine*, 7(74), 86–90 [in Ukrainian].
5. Sonnikova, V. I., Kulidzhoglyan, K. O. (2013). Model' kreditnogo skoringa kak algoritm tipologii nechetkih sovokupnostej [Model of credit scoring as algorithm of typology of fuzzy aggregations]. *Vestnik NGUJeU – Vestnik NSUEM*, 3, 107–117 [in Russian].
6. Ulanov, S. V. (2009). Ocenka kachestva i sravnenie skoringovykh kart [Evaluation of the scoring card quality]. *Jekonomicheskie nauki – Economic sciences*, 9(58), 330–335 [in Russian].
7. Abdou, H. A., Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent Systems in Accounting, Finance and Management Intell. Sys. Acc. Fin. Mgmt.*, 18(2–3), 59–88 [in English].
8. Anderson, R. (2007). *The credit scoring toolkit: theory and practice for retail credit risk management and decision automation*. Oxford: Oxford University Press [in English].
9. Hand, D. J., Henley, W. E. (1997). Statistical Classification Methods in Consumer Credit Scoring: A Review. *J Royal Statistical Soc A Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 523–541 [in English].
10. Liu, Y. (2001). New issues in credit scoring application. *Research paper 16/2001*. Institute of Information Systems, University of Goettingen [in English].
11. Liu, Y. (2002). A framework of data mining application process for credit scoring. *Research paper 01/2002*. Institute of Information Systems, University of Goettingen [in English].
12. Liu, Y. (2002). The evaluation of classification models for credit scoring. *Research paper 02/2002*. Institute of Information Systems, University of Goettingen [in English].
13. Lewis, E.M. (1992). *An introduction to credit scoring*. San Rafael, CA: Athena Press [in English].
14. Mays, E. (Ed). (2001). *Handbook of credit scoring*. Chicago: Glenlake Pub [in English].
15. Siddiqi, N. (2006). *Credit risk scorecards: developing and implementing intelligent credit scoring*. Hoboken, NJ: Wiley [in English].
16. Sohn, S.Y., Kim, D.H., & Yoon, J.H. (2016). Technology credit scoring model with fuzzy logistic regression. *Applied Soft Computing*, 43, 150–158 [in English].
17. Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16(2), 149–172 [in English].
18. Thomas, L. C., Crook, J., & Edelman, D. (2002). *Credit scoring & its applications – Siam Monographs on Mathematical Modeling and Computation*. Siam [in English].
19. Kebebew, E. (2006). Predictors of single-gland vs multigland parathyroid disease in primary hyperparathyroidism. *Arch Surg Archives of Surgery*, 141(8), 777–782 [in English].
20. Milchakov, K.S., Shebalkov, M.P. (2015). Skoringovyte karty v medicine: obzor i analiz publikacij [Scorecards in medicine: analytic review]. *Vrach i informacionnye tehnologii – Doctor and information technology*, 1, 71–79 [in Russian].
21. Malthouse, E.C. (1999). Ridge regression and direct marketing scoring models. *Journal of Interactive Marketing*, 13(4), 10–23 [in English].
22. Malthouse, E.C. (2001). Assessing the performance of direct marketing scoring models. *Journal of Interactive Marketing*, 15(1), 49–62 [in English].