

*І. В. Мірошниченко,  
к. е. н., доцент кафедри математичного моделювання та статистики,  
Київський національний економічний університет імені Вадима Гетьмана  
ORCID ID: 0000-0002-1307-7889*

*В. К. Крупін,  
магістрант спеціальності "Економічна кібернетика та Дата Сайанс",  
Київський національний економічний університет імені Вадима Гетьмана  
ORCID ID: 0000-0002-1001-6933*

DOI: 10.32702/2306-6814.2022.4.86

## ПРОГНОЗУВАННЯ БАНКРУТСТВА ПІДПРИЄМСТВА ЗА ДОПОМОГОЮ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ

*I. Miroshnychenko,  
PhD in Economics, Associate Professor of the Department of Mathematical Modeling and Statistics,  
Kyiv National Economic University named after Vadym Hetman*

*V. Krupin,  
Master's student of "Economic cybernetics and Data science",  
Kyiv National Economic University named after Vadym Hetman*

### BANKRUPTCY PREDICTION OF THE ENTERPRISE USING MACHINE LEARNING ALGORITHMS

**Статтю присвячено застосуванню методів машинного навчання для побудови моделей для прогнозування ймовірності банкрутства підприємства.**

**Встановлено доцільність використання алгоритмів машинного навчання для прогнозування банкрутства підприємств. Високу ефективність показали алгоритми Logistic Regression, Random Forest, XGBoost.**

**Обґрунтовано, що у сучасному світі керівники та менеджери організацій повинні приділяти велику увагу застосуванню машинного навчання у процесі визначення фінансового стану підприємств, оскільки завчасне виявлення кризового стану та застосування відповідних заходів для їх подолання допоможе, з одної сторони, підвищити інвестиційну привабливість компанії, з іншої — зміцнити соціально-економічне становище держави, а саме: зменшити кількість втрачених робочих місць, знизити рівень безробіття, відповідно збільшити податкові надходження до бюджетів всіх рівнів, а найголовніше — зменшити соціальну напругу у суспільстві.**

**The article is devoted to the application of machine learning methods to build models for predicting the probability of bankruptcy.**

**Determining the risk of bankruptcy of enterprises is one of the main problems of modern economic and financial research. Now, with increasing financial globalization, faster economic change, and a new dimension of increased financial risk in the context of a pandemic, the focus needs to be on improving the accuracy of the forecasting model and extending the forecasting horizon.**

**The consequences of financial failure are huge for financial lenders, managers, shareholders, investors, employees and even for the country's economy.**

**Accurate bankruptcy prediction usually leads to many benefits, such as reduced credit analysis costs, better monitoring, and so on. Thus, predicting bankruptcy becomes much more important. Today, the question is not whether to use bankruptcy forecasting models, but how to increase the efficiency of forecasting models.**

*The expediency of using machine learning algorithms to predict the bankruptcy of enterprises has been established. Algorithms Logistic Regression, Random Forest, XGBoost showed high efficiency. Python programming language is used to build machine learning models.*

*It is argued that in today's world, leaders and managers of organizations should pay close attention to the use of machine learning in determining the financial condition of enterprises, as early detection of crises and appropriate measures to overcome them will help on the one hand to increase investment attractiveness, on the other economic situation of the state, namely to reduce the number of lost jobs, reduce unemployment, respectively, increase tax revenues to budgets of all levels, and most importantly reduce social tensions in society.*

*It is noted that in order to increase the efficiency of forecasting the bankruptcy of enterprises, it is necessary to use the most balanced sample with approximately the same distribution of classes. Particular attention should be paid to the study of the representativeness of predictors, to determine their effectiveness for use in the model. As an alternative to classical approaches, models based on deep learning algorithms should be used, or an ensemble of different models should be used.*

*Ключові слова: банкрутство, машинне навчання, ризик, підприємництво, інвестиції.  
Key words: bankruptcy, machine learning, risk, entrepreneurship, investment.*

**ПОСТАНОВКА ПРОБЛЕМИ**

Визначення ризику банкрутства підприємств є однією з головних проблем сучасних економічних та фінансових досліджень. Нині з посиленням фінансової глобалізації, швидшими економічними змінами та новим виміром підвищеного фінансового ризику в контексті пандемії потрібно зосередитись на підвищенні точності моделі прогнозування та на продовженні горизонту прогнозування.

Наслідки фінансового провалу величезні для фінансових кредиторів, менеджерів, акціонерів, інвесторів, службовців і навіть для економіки країни. Саме тому протягом останніх п'яти десятиліть прогнозування банкрутства підприємств стало серйозною проблемою для різних зацікавлених сторін фірм. Точне прогнозування банкрутства зазвичай призводить до багатьох переваг, таких як зменшення витрат при кредитному аналізі, кращий моніторинг тощо. Таким чином, прогнозування банкрутства набуває набагато більшого значення. Сьогодні питання полягає не в тому, чи слід використовувати моделі прогнозування банкрутства, а в тому, як підвищити ефективність моделей прогнозування.

**АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ**

Графік ілюструє, що протягом 1968—1999 років було опубліковано дуже мало статей.

Можна помітити, що спостерігається зростаюча тенденція щодо кількості статей, опублікованих з 2000 року. Крім того, після світової фінансової кризи 2008 року дослідники почали набагато більше досліджувати цю область.

Відповідно до Закону України "Про відновлення платоспроможності боржника або визнання його банкрутом" банкрутство — це визнана господарським судом неспроможність боржника відновити свою платоспроможність за допомогою процедур санації та мирової угоди і погасити встановлені у порядку, визначеному цим Законом, грошові вимоги кредиторів не інакше як через застосування ліквідаційної процедури. Тобто законодавці ототожнюють поняття банкрутства передусім з процедурою ліквідації.

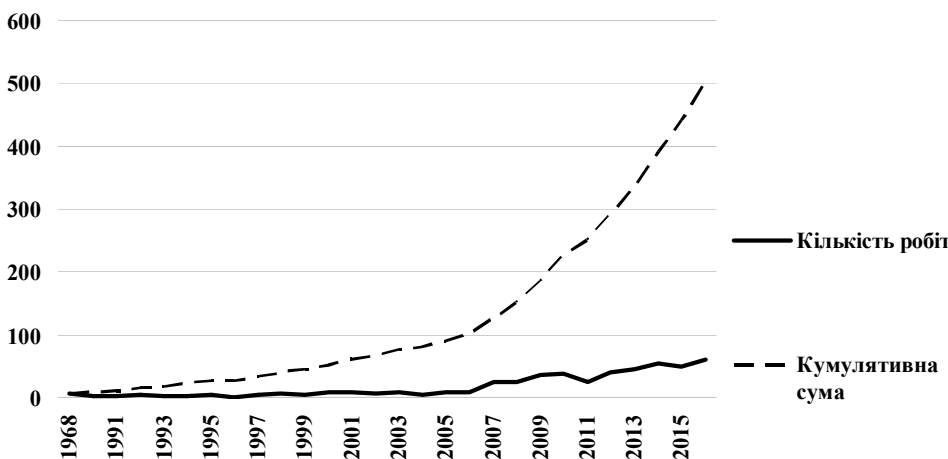
У науковій літературі також зустрічаються різні визначення банкрутства підприємства.

На думку О.О. Терещенка, під банкрутством розуміється ситуація, коли юридична особа неспроможна виконати зобов'язання перед бюджетом та задовольнити вимоги кредиторів [4, с. 23].

Схоже визначення надає Є.М. Андрущак, який стверджує, що головною ознакою банкрутства є неможливість розрахунків через нестачу активів у ліквідній формі, а тому такий незадовільний стан фізичної чи юридичної особи вважається банкрутством [1, с. 51].

Слід розуміти, що банкрутство є результатом кризи підприємства, після якої не-

Кількість робіт на тему прогнозування банкрутства, шт.



**Рис. 1. Зростання кількості робіт на тему банкрутства протягом 1968—2017 років**

можливо нормально провадити підприємницьку діяльність, що в свою чергу призводить до неплатоспроможності, а в подальшому до ліквідації підприємства. На думку Л.С. Ситник, банкрутство — це зафіксована юридично точка в життєвому циклі підприємства, що означає неможливість ведення фінансово-господарської діяльності з невід'ємним прибутком як самостійним господарюючим суб'єктом [4, с. 24].

Центральним теоретичним питанням про банкрутство підприємств є те, як пріоритетні правила впливають на ефективність рішень, прийнятих менеджерами (які, як передбачається, представляють інтереси власного капіталу), зокрема, чи інвестує фірма в безпечні чи ризиковані проєкти та чи подає заяву про банкрутство і коли. Неefективні інвестиційні рішення знижують прибуток фірми, що в свою чергу може призвести до незадовільного фінансового стану підприємства.

У західній літературі порівнюється ефективність корпоративних інвестиційних рішень, коли правило пріоритету при банкрутстві є правилом абсолютного пріоритету (ліквідація при банкрутстві), з тими, коли мають місце відхилення від правила абсолютного пріоритету (реорганізацію в умовах банкрутства). Загальновідомим результатом у фінансах є те, що капіталіст віддає перевагу ризикованим інвестиціям, аніж безпечним інвестиційним проєктам, оскільки отримує непропорційно високу вигоду, коли ризиковані проєкти досягають успіху, і несе лише обмежені збитки, коли ризиковані проєкти провалюються. Якщо правило пріоритету при банкрутстві змінюється на відхилення від правила абсолютного пріоритету, то перевага акціонерного капіталу до ризикованих проєктів стає ще сильнішою. Це пояснюється тим, що власний капітал тепер отримує позитивну віддачу, коли ризиковані проєкти провалюються, і таку саму високу віддачу, коли ризиковані проєкти досягають успіху [7, с. 95].

Методи прогнозування банкрутства фінансових фірм стали важливим питанням у 1960-х роках і широко досліджувались з тих пір.

Створення надійних моделей для прогнозування банкрутства є вирішальним для багатьох процесів прийняття рішень (Уеніш та Тон, 2017). Підхід, що використовується для прогнозування банкрутства, еволюціонував з часом починаючи з моделі Бівера (1966, 1968), заснованої на одновимірному аналізі для вибраних співвідношень, яка мала дуже гарну прогнозовану силу. Потім Альтман (1968) досяг успіху, розробивши модель множинного дискримінантного аналізу, яка називається моделлю Z-Score. У літературі про передбачення банкрутства найбільш цитованими є моделі Альтмана (1968), Олсона (1980) та Змієвського (1984), які базуються на облікових змінних. Ці моделі прогнозування банкрутства використовують різні пояснювальні змінні та статистичні методи. Тому прогнозована сила цих моделей прогнозування банкрутства відрізняється. Однак, коли використовуються оригінальні статистичні методи, показники точності для моделей Альтмана (1968), Олсона (1980) та Змієвського (1984) становлять відповідно 80,6%, 93,8% та 95,3% [10, с. 52]. Вивчаючи ефективність моделі z-score Альтмана при прогнозуванні банкрутства спеціалізованих роздрібних фірм, що ве-

дуть бізнес у сучасні часи, Чайтанья (2005) виявив, що всі банкрутства, крім двох (94%), були б точно передбачені. У 1970-х роках популярними стали методи logit та probit. Потім, у 1990-х роках, стали застосовувати штучні нейронні мережі та генетичні алгоритми [19, с. 19].

Моделі прогнозування банкрутства можна розділити на дві загальні категорії залежно від типу використовуваної змінної: статичні моделі (Альтман 1968, 2000, 2002; Тафлер 1982, 1983, 1984; Ольсон 1980; Змієвський 1984; Теодосіо 1991) або динамічні моделі (Шумвей 2001; Гілеест 2004). Хоча статистичні методи стали найбільш часто використовуваними при прогнозуванні банкрутства, вони характеризуються багатьма недоліками щодо статистичних припущень, таких як лінійність, нормальність та незалежність серед змінних, які були виявлені в багатьох дослідженнях (наприклад, Альтман 2018; Баклаен 2006; Северін 2011; Джаясекера 2018). Ось чому за останні два десятиліття популярність методів прогнозування банкрутства перейшла від статистичних до інтелектуальних [21, с. 167]. У 21 ст. поряд з розвитком інформаційних технологій були розроблені різноманітні методики, що використовуються для прогнозування банкрутства. Основний акцент зосереджений на машинах векторної підтримки, нечіткій логіці, випадкових лісах та підходах з кількома моделями [9, с. 4530].

За останнє десятиліття в ряді досліджень застосовувались штучні інтелектуальні методи для прогнозування банкрутства. В даний час ці методи включають дерева рішень, різні архітектури штучної нейронної мережі (ANN), включаючи багатошарове сприйняття (MLP), самоорганізуючу карту (SOM) та квантування векторів навчання (LVQ), еволюційні підходи, включаючи генетичні алгоритми, інші інтелектуальні методи, включаючи машини векторної підтримки (SVM) [5, с. 28].

Ключовим недоліком у моделях прогнозування банкрутства є проблема їх застарівання з плином часу з часу їх розробки. У літературі передбачається, що моделі прогнозування ризику банкрутства працюють добре протягом чотирьох-шести років, після чого їх необхідно модифікувати. Однак слід зазначити, що життєвий цикл моделі досить мінливе поняття. Немає суворих правил, які точно визначають, коли закінчується життя моделі. З цього приводу слід продемонструвати здоровий глузд. Моделі можуть втратити свою актуальність через зміни у діловому циклі та зміни в економічних умовах, через що середні значення економічних та фінансових показників можуть змінюватися. Додавання динамічної перспективи до моделі може збільшити термін дії моделей.

Отже, багато сучасних методів прогнозування банкрутства необхідно постійно вдосконалювати, враховувати не тільки набір певних фінансових показників, а також вплив часової складової тощо.

## МЕТА СТАТТІ

Метою роботи є застосування методів машинного навчання для побудови моделі, яка на основі розноманітної інформації буде прогнозувати ймовірність банкрутства підприємства.

## ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

До основних цілей роботи слід віднести акцентування на важливості прогнозування банкрутства; дослідження методів, що використовуються для прогнозування; доведення переваг застосування машинного навчання для обробки великого обсягу інформації; порівняння альтернативних моделей, що використовуються для вирішення задач класифікації; проведення економічних висновків на основі отриманих результатів.

## ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ ДОСЛІДЖЕННЯ

Сьогодні банкрутство визнається загальним явищем. Банкрутство — це реалізація катастрофічних ризиків компанії в процесі її діяльності, що призводить до невиконання строкових вимог кредиторів та виконання бюджетних зобов'язань. Основними причинами цього явища є такі:

- значне недотримання фінансової стійкості компаній у разі перевищення зобов'язань над активами;
- тривала неплатоспроможність компанії;
- низька ліквідність активів;
- значний дисбаланс позитивних та негативних грошових потоків [3, с. 683].

Банкрутство в ринковій економіці служить регулятором, який "відкидає" збитковий, неефективний та неплатоспроможний бізнес. Проте банкрутство спричиняє низку негативних соціально-економічних наслідків, серед яких найважливішим є поширення безробіття в країні.

Зростаюча кількість ліквідованих та збанкрутілих компаній в Україні спонукає до пошуку шляхів вирішення цього питання на національному рівні, що у свою чергу створює потребу у вивченні основних проблем та причин банкрутства українських компаній. Отже, питання банкрутства підприємств в Україні, безперечно, є актуальним.

За останні п'ять років багато технологій перейшли від концепцій до реальності. Вони створили нові захоплюючі можливості для корпорацій у процесі. Такі технології, як штучний інтелект (ШІ), машинне навчання (МН) та можливості аналізу даних, тепер вбудовані в практичні програми для покращення бізнес-аналітики (BI), розширення підтримки прийняття рішень, забезпечення більшої та швидшої ефективності обробки, швидшого виявлення нових вразливостей, і мінімізації витрат та підвищення прибутковості. Концепції машинного навчання перетворились із неясних ідей, про які багато хто з нас чув лише у науково-фантастичних фільмах, на те, щоб відіграв важливу роль у повсякденному житті пересічної людини, додатки на основі машинного навчання стали загальноприйнятною концепцією програмного забезпечення, і його ефект буде ставати все більш помітним та поширеним із роками [8, с. 280].

Для побудови моделей машинного навчання були зібрані дані з інформаційної служби нових ринків EMIS, яка є базою даних, що містить інформацію про ринки, що розвиваються по всьому світу. Компанії-банкрути були проаналізовані в період 2000—2012 рр. Водночас підприємства, що продовжують функціонувати, оцінювалися з 2007 по 2013 роки [20].

На основі зібраних даних було виділено п'ять випадків класифікації, що залежить від періоду прогнозування:

— 1-й рік — дані містять фінансові ставки з 1-го року періоду прогнозування та відповідне значення класу, що вказує на стан банкрутства через 5 років. Дані містять 7027 випадків (фінансових звітів), 271 представляє збанкрутовані компанії, 6756 фірм, які не збанкрутували в прогнозний період.

— 2-й рік — дані містять фінансові ставки з 2-го року періоду прогнозування та відповідне значення класу, що вказує на стан банкрутства через 4 роки. Дані містять 10173 випадки (фінансові звіти), 400 представляють збанкрутовані компанії, 9773 фірми, які не збанкрутували в прогнозний період.

— 3-й рік — дані містять фінансові ставки з 3-го року періоду прогнозування та відповідне значення класу, що вказує на стан банкрутства через 3 роки. Дані містять 10503 випадки (фінансові звіти), 495 представляють збанкрутовані компанії, 10008 фірм, які не збанкрутували в прогнозний період.

— 4-й рік — дані містять фінансові ставки з 4-го року періоду прогнозування та відповідне значення класу, що вказує на стан банкрутства через 2 роки. Дані містять 9792 випадки (фінансові звіти), 515 представляють збанкрутовані компанії, 9277 фірм, які не збанкрутували в прогнозний період.

— 5-й рік — дані містять фінансові ставки з 5-го року періоду прогнозування та відповідне значення класу, що вказує на стан банкрутства після 1 року. Дані містять 5910 випадків (фінансових звітів), 410 представляє збанкрутовані компанії, 5500 фірм, які не збанкрутували в прогнозний період.

Спочатку потрібно об'єднати дані з п'яти таблиць, щоб отримати один датасет.

В результаті отримано датасет, який складається з 65 колонок та 43405 рядків з записами.

Набір даних є незбалансованим, оскільки 41314 спостережень відноситься до класу 0 і лише 2091 спостережень до класу 1.

Перед побудовою моделей потрібно виконати етап обробки даних.

### 1. Заповнення пропусків.

Найпростіший варіант полягає у виключенні об'єктів, що мають неповні відомості (тобто видалення тих рядків з матриці ознак, в шпальтах яких є пропуски), або виключення ознак, що містять неповні відомості (тобто видалення тих стовпців, в яких є пропуски). Плюси цього варіанту в тому, що він дуже простий і його можна відразу випробувати на будь-якій моделі. Мінуси цілком очевидні. Якщо багато об'єктів будуть мати пропуски, то ми ризикуємо видалити важливі аспекти закономірності, яка прихована в даних, і, як наслідок, отримаємо низьку якість апроксимації. Якщо ми видалимо стовпець, який містить вкрай важлива ознака, ми ризикуємо зовсім втратити ключову інформацію про розділення об'єктів. Другий варіант боротьби з пропусками полягає в тому, щоб замінити їх за допомогою інтерполяції. Це може бути середнє або медіанне значення по стовпцю. У разі якщо ознака є функцією від часу, як і окремі об'єкти, то можна інтерполювати пропуски тільки по сусідніх тимчасовим значенням (наприклад, якщо об'єкт є ймовірність покупки квартири, а ознака — середню заробітну плату програміста в цей рік, то за пропущений рік цифру можна наблизити по двом сусіднім) [6, с. 44].

**Таблиця 1. Показники ефективності побудованих моделей**

Метрика	Logistic Regression	Random Forest	XGBoost
Accuracy	0.699	0.968	0.981
Precision	0.106	0.880	0.843
Recall	0.683	0.416	0.769
ROC_AUC	0.691	0.707	0.881

Атрибути, в яких пропущено 50% даних, будуть видалені, оскільки не доцільно використовуватися їх у якості вагомих предикторів (такою змінною виявилася Attr37), інші пропуски будуть заповнені середнім значенням (при чому для кожного класу окремо, щоб не змінити розподіл даних для кожного класу).

**2. Видалення викидів.**

Значення, що є досить великими (більше ніж 3 стандартні відхилення від середнього), є аномальними, не несуть статистично важливою інформації і відповідно погіршують побудовані моделі.

Одним з найпростіших методів виявлення викидів є використання *boxplot* — це графік для опису розподілу даних. Графічні ділянки використовують медіану та нижній та верхній квантілі.

Також іншим підходом є видалення викидів за допомогою допомогою *Z-score*.

$$z_i = \frac{x_i - \bar{x}}{\sigma} \tag{1}$$

де  $x_i$  — спостереження по фактору;

$\bar{x}$  — середнє значення;

$\sigma$  — стандартне відхилення.

Викидами вважаються значення, що відхиляються від середнього значення більше ніж на 3 стандартні відхилення.

Рядки датасету, для яких значення є більшими за допустиму норму були видалені (для кожного класу ця процедура була виконана окремо).

В результаті залишилося 40259 спостережень класу 0 та 1924 спостереження класу 1.

**3. Перевірка на мультиколінеарність (наявність лінійної залежності між двома або більше факторними змінними).**

Перевірка на мультиколінеарність є важливою, оскільки її наслідками є:

- зміщення оцінок параметрів моделі;
- збільшення коваріації оцінок;
- незначущість параметрів моделі [6, с. 48].

Для кожної пари факторів було розраховано коефіцієнт кореляції Пірсона.

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{2}$$

де  $x_i$  — спостереження по фактору 1;

$\bar{x}$  — середнє значення фактору 1;

$y_i$  — спостереження по фактору 2;

$\bar{y}$  — середнє значення фактору 2.

У результаті побудови матриці парної кореляції, змінні які мають кореляцію більшу за 0,9 (тісний зв'язок) були видалені.

У результаті було видалено 16 змінних, а загальна кількість факторів зменшилася до 47.

**4. Проведено нормалізацію показників.**

Необхідність нормалізації вибірок даних зумовлена тим, що значення факторів можуть змінюватися в дуже великому діапазоні і відрізнятися один від одного на кілька порядків. Для машинного навчання кожен набір даних не потребує нормалізації. Це потрібно лише тоді, коли змінні мають різний діапазон.

Нормалізація *min-max* — один із найпоширеніших способів нормалізації даних. Для кожної функції мінімальне значення цієї функції перетворюється на 0, максимальне значення перетворюється на 1, а кожне інше значення перетворюється на десяткове число від 0 до 1 [12].

$$x_n = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{3}$$

де  $x_i$  — спостереження по фактору.

$x_{min}$  — мінімальне значення.

$x_{max}$  — максимальне значення.

**5. Розбиття датасету на тренувальну та тестувальну вибірку.**

Необхідним етапом після підготовки даних є їх розподіл на навчальну та тестову вибірку.

Модель будується за допомогою навчального набору даних, який є множиною прикладів, які використовуються для налаштування параметрів моделі. На підставі різниці між реальними та спрогнозованими значеннями цільового показника параметри моделі коригуються. Далі натренована модель використовується для прогнозування цільової змінної у другому наборі даних, що називається тестовим набором.

Вибірку розбито на тренувальну та тестову у співвідношенні 75% та 25%.

Крім того, як зазначалося, датасет є незбалансованим, саме тому було виконано процедуру семплювання для тренувальної вибірки, тобто кількість спостережень класу 1 дубльовано випадковим чином так, щоб їх кількість дорівнювала кількості спостережень класу 0.

У результаті тренувальна вибірка містить 60474 спостережень (30237 класу 0 та 30237 класу 1), тестова вибірка містить 10546 спостережень (10022 класу 0 та 524 класу 1).

Було побудовано 3 моделі (Logistic Regression, Random Forest, XGBoost). Моделі порівняні за допомогою таких показників:

— Accuracy — вимірює точність моделі класифікації як частку справжніх результатів у загальній кількості випадків.

— Recall — це частка всіх правильних результатів, які повертає модель.

— Precision — це частка справжніх результатів над усіма позитивними результатами.

— ROC AUC — співвідношення між Recall та Specificity (TN/(TN+FP)).

За кількістю правильно визначених випадків банкрутства найкращою моделлю є XGBoost (TP=403, FP=121), значення Accuracy, Recall, ROC\_AUC також є найвищими, тобто цей алгоритм найкраще підходить для вирішення цієї задачі. Найгірші показники має класична модель — Logistic Regression. Random Forest трохи точніше визначає фірми, що не збанкрутують, але має найгірші показники щодо випадків банкрутства.

На основі найкращої моделі можна визначити найбільш вагомі фінансові показники, на які потрібно орієнтуватись для запобігання банкрутства.

Так, найбільш вагомими є саме такі показники:

1. (валовий прибуток + амортизація) / сукупні зобов'язання

Чим більша сума валового прибутку та амортизації та чим менші сукупні зобов'язання, тим більше ймовірність фірми не зазнати фінансової невдачі.

2. Операційні витрати / загальні зобов'язання

Успішними є компанії з вищими операційними витратами та нижчими загальними зобов'язаннями.

3. Прибуток від операційної діяльності / фінансові витрати

Чим більший прибуток від операційної діяльності та чим менші фінансові витрати, тим більше ймовірність фірми не зазнати фінансової невдачі.

4. Обсяг продажів (n) / обсяг продажів (n-1)

Цей показник суттєво відрізняється для підприємств, які функціонують 1—2 роки.

5) нерозподілений прибуток / сумарні активи

Для ліквідованих підприємств цей показник взагалі має суто від'ємні значення, тобто такі фірми мали нерозподілені збитки, що у свою чергу призвело до банкрутства.

Отже, алгоритм XGBoost та процес машинного навчання дають змогу в 77% випадків визначити, що фірма має незадовільний фінансовий стан та знаходиться на межі банкрутства, що, на мою думку, для такої незбалансованої вибірки є досить якісним результатом, тим паче, змінюючи порогове значення, ми отримуємо результат, який буде більш бажаним для нас. Крім того, ми маємо можливість визначити найбільш вагомі фінансові показники, на яких треба зацентувати увагу, щоб фірма не зазнала фінансової невдачі. Вважаємо, що з точки зору бізнесу такі моделі є дуже корисними, оскільки:

1. Дають більше розуміння власникам бізнесу щодо поточного стану підприємства та дозволяють приймати рішення не тільки на основі фінансових звітів.

2. Є корисними для інвесторів, які не можуть вирішити, чи є сенс вкладати кошти в це підприємство. Якщо ймовірність фірми зазнати фінансового провалу дуже низька, то це буде залучати більше інвесторів, оскільки у них буде впевненість у тому, що вони не втратять свої кошти.

3. Можуть бути корисними для уряду країни, оскільки знаючи, яка частка підприємств в якій галузі знаходиться на межі фінансового провалу, можна допомагати галузям, які мають найбільші проблеми.

Таким чином, математичні моделі та машинне навчання при правильному використанні можуть стати дуже потужним апаратом, який за короткий час та за мінімальної затрати ресурсів дозволяє отримати максимальну вигоду для суб'єктів економічної діяльності.

## ВИСНОВКИ

У роботі було використано алгоритми Logistic Regression, Random Forest, XGBoost для побудови моделей, метою яких є пошук оптимального рішення задачі класифікації: прогнозування банкрутства. На основі отриманих результатів можна зробити висновок, що використання машинного навчання є доцільним методом вирішення таких задач, оскільки було отримано

результат у 77% правильно спрогнозованих значень банкрутства (дані є незбалансованими, оскільки лише 4.8% підприємств є банкрутами), а загальна точність моделі дорівнює 98%, при цьому процес навчання займає декілька хвилин, що доводить переваги у порівнянні з застосуванням експертів та людського інтелекту.

На нашу думку, у сучасному світі керівники та менеджери організацій повинні приділяти велику увагу застосуванню машинного навчання у процесі визначення фінансового стану підприємств, оскільки завчасне виявлення кризового стану та застосування відповідних заходів для їх подолання допоможе зміцнити соціально-економічне становище держави, а саме зменшити кількість втрачених робочих місць, знизити рівень безробіття, відповідно збільшити податкові надходження до бюджетів всіх рівнів, а найголовніше зменшити соціальну напругу у суспільстві.

Серед основних пропозицій для подальшого дослідження можна запропонувати таке:

— для більшої точності необхідно використовувати більш збалансовану вибірку з приблизно однаковим розподілом класів;

— окрему увагу слід приділити дослідженню репрезентативності предикторів, визначити їх ефективність для використання у моделі;

— використовувати не тільки кількісні показники, а й якісні (наприклад, сферу діяльності підприємства).

Однією з альтернатив зазначеним підходам, можна використати моделі, що засновані на алгоритмах глибокого навчання, або використовувати ансамбль з різних моделей.

## Література:

- Андрущак ЄМ. Удосконалення інституту банкрутства. Фінанси України, Київ, 2006. № 9. С. 50—56.
- Івченко К.В. Машинне навчання та когнітивні обчислення. Матеріали міжнародної науково-практичної інтернет-конференції "Використання інноваційних технологій в процесі підготовки фахівців", Вінниця, 2017. С. 1—4.
- Сукрушева Г.О., Папуцин В.М. Причини та наслідки банкрутства вітчизняних підприємств у сучасних умовах. Економіка та суспільство: електронне наук. фах. вид. Мукачево: МДУ, 2017. С. 682—686.
- Терещенко О.О. Фінансова санація та банкрутство підприємств: навч. посіб. Київ: КНЕУ, 2011. 412 с.
- Тимошук О.Л., Дорундяк К.М. Оцінювання ймовірності банкрутства підприємств за допомогою дискримінантного аналізу та нейронних мереж. Системні дослідження та інформаційні технології. Київ. 2018. С. 22—34.
- Трофименко В.В. Дослідження використання методів машинного навчання в мережах інтернет речей. Київ: КПП, 2019. 87 с.
- Aleksanyan L., Huiban J.P. Economic and financial determinants of firm bankruptcy: evidence from the French food industry. Rev Agric Food Environ Stud 97, 2016. Pp. 89—108.
- Attaran M. and Deb P. 'Machine learning: the new 'big thing' for competitive advantage', Int. J. Knowledge Engineering and Data Mining, Vol. 5. № 4. 201. Pp. 277—305.
- Chen M.Y. Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of

evolutionary computation. *Comput. Math. Appl.* 62 (12), 2011. Pp. 4514—4524.

10. Dagmar C., Jiri K. Comparison of Prediction Models Applied in Economic Recession and Expansion Reprinted from: *J. Risk Financ. Manag.* 2020. Pp.46—60.

11. Deb P. The Logistic Regression Algorithm. URL: <https://machinelearning-blog.com/2018/04/23/logistic-regression-101/> (Accessed: 30.11.2021).

12. Everything You Need to Know About Classification in Machine Learning. URL: <https://www.simplilearn.com/tutorials/machine-learning-tutorial/classification-in-machine-learning> (Accessed: 30.11.2021).

13. Fred Weston J.F., Some Economic Fundamentals for an Analysis of Bankruptcy. 41 *Law and Contemporary Problems*. 2007. Pp. 47—65.

14. Grybinenko O. Social and economic consequences of bankruptcy of the companies in Ukraine: "EUREKA: Social and Humanities". Number 2. 2017. Pp. 3—10.

15. Hong S. Ensemble — Bagging and Boosting. URL: <https://seongjuhong.com/2021-01-17pm-ensemble-bagging-and-boosting/> (Accessed: 30.11.2021).

16. Mbaabu O. Introduction to Random Forest in Machine Learning URL: <https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/> (Accessed: 30.11.2021).

17. Michelle J. White. "Corporate and Personal Bankruptcy Law," NBER Working Papers 17237, National Bureau of Economic Research, Inc, 2011. Pp. 1—9.

18. Morde V. XGBoost Algorithm: Long May She Reign! URL: <https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d> (Accessed: 30.11.2021).

19. Nicoleta B., Mara. Assessment of Bankruptcy Risk of Large Companies: European Countries Evolution Analysis Reprinted from: *J. Risk Financ. Manag.* 2020. Pp. 17—45.

20. Polish companies bankruptcy data. URL: <https://archive.ics.uci.edu/ml/datasets/Polish+companies+bankruptcy+data> (Accessed: 30.11.2021).

21. Tomasz K. Dynamic Bankruptcy Prediction Models for European Enterprises Reprinted from: *J. Risk Financ. Manag.* 2019. Pp. 155—171.

#### References:

1. Andruschak, Y. M. (2006), "Improving the institution of bankruptcy", *Finance Ukraine*, vol. 9, pp. 50—56.

2. Ivchenko, K. V. (2017), "Machine learning and cognitive calculations", *Materialy mizhnarodnoi naukovo-praktychnoi internet-konferentsii. Vykorystannia innovatsijnykh tekhnolohij v protsesi pidhotovky fakhivtsiv* [Proceedings of the international scientific-practical Internet conference. The use of innovative technologies in the training of specialists], Vinnytsia, Ukraine, pp. 1—4.

3. Sukrusheva, H. O. and Paputsyn, V. M. (2017), "Causes and consequences of bankruptcy of domestic enterprises in modern conditions", *Economics and society*, *Mukachiv: derzh. un-t. Mukachevo*, pp. 682—686.

4. Tereschenko, O.O. (2011), *Finansova sanatsiia ta bankrutstvo pidprijemstv* [Financial rehabilitation and bankruptcy of enterprises], KNEU, Kyiv, Ukraine.

5. Tymoschuk, O. L. and Dorundiak, K. M. (2018), "Estimation of the probability of bankruptcy of enterprises

using discriminant analysis and neural networks", *Systemni doslidzhennia ta informatsijni tekhnolohii* [System research and information], Kyiv, Ukraine, pp. 22—34.

6. Trofymenko, V.V. (2019), *Doslidzhennia vykorystannia metodiv mashynnoho navchannia v merezhakh internet rechej*, [Research of the use of machine learning methods in the Internet of Things], KPI, Kyiv, Ukraine.

7. Aleksanyan, L. and Huiban, J.P. (2016), "Economic and financial determinants of firm bankruptcy: evidence from the French food industry", *Rev Agric Food Environ*, vol. 97, pp. 89—108.

8. Attaran, M. and Deb, P. (2018), "Machine learning: the new 'big thing' for competitive advantage", *Int. J. Knowledge Engineering and Data Mining*, vol. 5, No. 4. pp. 277—305.

9. Chen, M.Y. (2011), "Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation", *Comput. Math. Appl.*, vol. 62 (12), pp. 4514—4524.

10. Dagmar, C. and Jiri, K. (2020), "Comparison of Prediction Models Applied in Economic Recession and Expansion", *Risk Financ. Manag.*, pp. 46—60.

11. Deb, P. (2021), "The Logistic Regression Algorithm", available at: <https://machinelearning-blog.com/2018/04/23/logistic-regression-101/> (Accessed: 30.11.2021).

12. Simpli Learn (2021), "Everything You Need to Know About Classification in Machine Learning", available at: <https://www.simplilearn.com/tutorials/machine-learning-tutorial/classification-in-machine-learning> (Accessed: 30.11.2021).

13. Fred Weston, J.F. (2007), "Some Economic Fundamentals for an Analysis of Bankruptcy", 41 *Law and Contemporary Problems*, pp. 47—65.

14. Grybinenko, O. (2017), "Social and economic consequences of bankruptcy of the companies in Ukraine", *EUREKA: Social and Humanities*, vol. 2, pp. 3—10.

15. Hong, S. (2021), "Ensemble — Bagging and Boosting", available at: <https://seongjuhong.com/2021-01-17pm-ensemble-bagging-and-boosting/> (Accessed: 30.11.2021).

16. Mbaabu, O. (2021), "Introduction to Random Forest in Machine Learning", available at: <https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/> (Accessed: 30.11.2021).

17. Michelle, J. White. (2011), "Corporate and Personal Bankruptcy Law", NBER Working Papers 17237, National Bureau of Economic Research, Inc, pp. 1—9.

18. Morde, V. (2021), "XGBoost Algorithm: Long May She Reign! ", available at: <https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d> (Accessed: 30.11.2021).

19. Mara, N. B. (2020), "Assessment of Bankruptcy Risk of Large Companies: European Countries", *Evolution Analysis* Reprinted from: *J. Risk Financ. Manag.*, pp.17—45.

20. UCI (2021), "Polish companies bankruptcy data", available at: <https://archive.ics.uci.edu/ml/datasets/Polish+companies+bankruptcy+data> (Accessed: 30.11.2021).

21. Tomasz, K. (2019), "Dynamic Bankruptcy Prediction Models for European Enterprises", Reprinted from: *J. Risk Financ. Manag.*, pp. 155—171.

*Стаття надійшла до редакції 07.02.2022 р.*