
**МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ
ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ЕКОНОМІЦІ**

УДК 330.46

DOI: <https://doi.org/10.32782/2520-2200/2021-6-13>**Пишнограєв І.А.**кандидат економічних наук, доцент
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»**Васильцова Ю.В.**студентка
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»**Pyshnograev Ivan**National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»**Vasiltsova Yuliia**National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»**МОДЕЛЮВАННЯ ВПЛИВУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ
КРИЗОВИХ ЯВИЩ НА МІГРАЦІЙНІ ПРОЦЕСИ****MODELING THE IMPACT OF SOCIO-ECONOMIC CRISES
ON MIGRATION PROCESS**

Міграційні рухи, зокрема зовнішня міграція, характерні для всіх держав світу, і Україна не виняток. Особливість ролі переміщень між державами полягає у можливості за їх допомогою досягти балансу між кількісним та якісним складом трудових ресурсів, тому дуже важливо розуміти сутність міграційного процесу та вміти його прогнозувати. За допомогою використання методів машинного навчання побудовано прогноз розвитку зовнішнього міграційного процесу для України з урахуванням впливу на нього економічних чинників та індикаторів кризових явищ. Як показник міграційного руху використано відносне міграційне сальдо, що розраховується як співвідношення між чисельністю іммігрантів на емігрантів. Для виконання роботи застосовано методи кореляційного та регресійного аналізу, теорії статистики, ансамблеві методи машинного навчання, методи прогнозування часових рядів та побудова нейронних мереж. Модель, представлена у статті, може бути використана для прогнозування характеру міграційного руху між державами з урахуванням впливу різних чинників.

Ключові слова: відносне міграційне сальдо, кореляційний аналіз, регресійний аналіз, часовий ряд, метод «випадкового лісу», метод зменшення градієнту, нейронна мережа, прогнозування.

Миграционные движения, в частности внешняя миграция, характерны для всех государств мира, и Украина не является исключением. Особенность роли перемещений между государствами заключается в возможности с их помощью достичь баланса между количественным и качественным составом трудовых ресурсов, поэтому очень важно понимать сущность миграционного процесса и уметь его прогнозировать. При помощи использования методов машинного обучения построен прогноз развития внешнего миграционного процесса для Украины с учетом влияния на него экономических факторов и индикаторов кризисных явлений. В качестве показателя миграционного передвижения было использовано относительное миграционное сальдо, которое рассчитывается как соотношение между численностью иммигрантов и эмигрантов. Для выполнения работы применены методы корреляционного и регрессионного анализа, теории статистики, ансамблевые методы машинного обучения, методы прогнозирования временных рядов и построение нейронных сетей. Модель, представленная в статье, может быть использована для прогнозирования характера миграционного движения между государствами с учетом влияния разных факторов.

Ключевые слова: относительное миграционное сальдо, корреляционный анализ, регрессионный анализ, часовой ряд, метод «случайного леса», метод уменьшения градиента, нейронная сеть, прогнозирование.

Migration movements, in particular external migration, are common to all countries of the world, and Ukraine is not an exception. The peculiarity of the role of movement between states is the ability to achieve a balance between the quantitative and qualitative composition of the labor force, so it is important to understand the nature of the migration process and be able to predict it for decision-making at the state level. Modelling of the external migration process are taking into account the influence of economic factors, so using of machine learning methods is proposed. This choice of methodology is explained by the fact that machine learning allows to obtain high values of accuracy in forecasting of socio-economic phenomena. Indicators of socio-economic crises, as well as factors characterizing the level of development of the state were used in the selection of factors influencing the external migration process. The migration process was predicted using the "random forest" method. The selection of the model was based on minimizing the deviations of the predicted data from the actual ones. The model was trained on statistical information from 15 countries, covering a period of 20 years. Countries were selected according to the size of migration flows with Ukraine, taking into account both the flow of departure and the flow of arrival. Net migration rate, counted by dividing the number of immigrants by the number of emigrants, was used as the indicator of migration movement. To predict the importance of influencing factors, time series forecasting methods were used, in particular the ARIMA model, the choice of the best model was based on minimizing the Akaike criterion. The obtained results were used to forecast the external migration process for Ukraine for 2020–2021. In the course of the work the methods of correlation and regression analysis, theories of statistics, ensemble methods of machine learning, methods of forecasting time series and construction of neural networks were used. The model presented in the article can be used to predict the nature of migration between states, taking into account the influence of various factors.

Key words: relative migration balance, correlation analysis, regression analysis, time series, random forest method, gradient reduction method, neural network, forecasting.

Постановка проблеми. Поширення процесів глобалізації призводить до посилення зв'язків між країнами. Взаємовідносини проявляються не лише в торгівлі між країнами, а й в обміні ресурсами, у тому числі трудовими, що реалізується за рахунок міграційних процесів. Ураховуючи вплив міграції на рівень соціально-економічного розвитку, важливим є розроблення інструментарію, що дасть змогу моделювати міграційний процес. Основна ціль моделей міграції – надати аналітичну структуру, за допомогою якої виявляються основні форми впливу на міграційний процес та прогнозується вплив зовнішніх чинників.

Аналіз останніх досліджень і публікацій. Найбільш ранні моделі міграції являли собою факторні моделі, що враховували чинники регіонів, між якими відбуваються міграційні переміщення. До них, зокрема, належать гравітаційні моделі міграції, що пояснюють вплив двох чинників: населення регіонів та відстані між ними. Гравітаційні моделі для опису механічного руху використовували у своїх працях російські вчені М.П. Кулаков, М.Ю. Хавінсон [1], В.А. Москвіна [2] Недоліком даного класу моделей є врахування недостатньої кількості факторів для опису міграційного руху, тому кращими є модифіковані гравітаційні моделі, відображені в працях Н.А. Трофимової, В.А. Розумовської, Л.В. Антосіка, Н.В. Івашиної [3] Достатньо новою є модель випромінювання, що прогнозує міграційні потоки, орієнтуючись також на фактори регіонів, що знаходяться між пунктами призначення та вибуття [5].

Застосування економетричних моделей для опису міграційного процесу є характерним для російських науковців О.І. Стебунової, А.С. Румянцева, Н.П. Маслова [6], Г.А. Батищевой, М.І. Журавльової [7]

Використання методів математичної статистики та багатофакторного аналізу є недостатньо ефективним, коли присутня висока зашумленість початкової вибірки. Також недоліком традиційних моделей є недостатня кількість даних, що оцінюють міграційний процес. Тому для аналізу міграції все частіше починають використовувати методи машинного навчання, зокрема побудову нейромереж, що дають змогу будувати складні нелінійні зв'язки між змінними та досягати високої точності.

Метою дослідження є моделювання міждержавного міграційного процесу з метою прогнозування інтенсивності зовнішньої міграції, ураховуючи вплив економічних чинників на механічний рух населення.

Виклад основного матеріалу дослідження. Міграція – тип мобільності, за якого людина під час переїзду має намір змінити своє поточне місце проживання. Максимально точне прогнозування міграційного процесу є дуже важливим для міського планування, міжнародної торгівлі, планування природоохоронного процесу та розроблення державної політики.

За поширення використання великих обсягів даних (big data) традиційним методам моделювання міграційного процесу не вдається виконувати поставлені завдання, тому все більшої популярності набувають методи машинного

навчання. Під машинним навчанням (machine learning) зазвичай розуміють групу математичних, статистичних та обчислювальних методів, що вирішують різні завдання шляхом визначення закономірностей між вхідними даними, тому машинне навчання використовується для прогнозування, прийняття рішень та діагностики в різних сферах [9].

Розрізняють два методи машинного навчання: класичні та некласичні. Сьогодні все більшої популярності набувають саме некласичні методи, до яких зазвичай відносять нейронні мережі та побудову генетичних алгоритмів.

У цій роботі для моделювання міграційного процесу були використані методи мови R, реалізовані в середовищі RStudio. Основною метою використання мови є виконання математичних обчислень, що пов'язані з машинними методами та ґрунтуються на статистичних даних. Тому в разі коли необхідно вирішувати задачі, що значною мірою засновані на статистиці, ця мова буде найбільш удалим вибором.

Сьогодні найбільш точні результати під час моделювання дають такі методи машинного навчання, як ансамблеві методи та побудова нейронних мереж. До основних способів побудови ансамблів відносять:

1. Стекінг: полягає у навчанні декількох різних алгоритмів на вхідних даних, у результаті отримується значення, яке найчастіше зустрічається в процесі моделювання. Порівняно з двома наступними методами стекінг характеризується найменшою точністю.

2. Беггінг (Bootstrap AGGregating) полягає у паралельному навчанні алгоритму на великій кількості випадкових вибірок із вихідної бази даних, а кінцевий результат отримується як середнє значення проміжних. До найбільш популярного алгоритму беггінгу відносять Random Forrest.

3. Бустінг полягає у послідовному навчанні алгоритмів на вибірках із вихідної бази даних, але дані для наступного етапу навчання вибираються не випадково, як у беггінгу, а ту частину даних, на якій алгоритм спрацював неправильно. Саме цей метод характеризується найбільшою точністю у вирішенні завдання класифікації.

Нейромережі використовуються для вирішення всіх завдань, що постають перед іншими методами машинного навчання. Будь-яка мережа являє собою набір нейронів та взаємозв'язків між ними. Даний метод являє собою функцію з багатьма входами та одним виходом.

Процес навчання під час використання machine learning ділиться на два етапи:

1. Етап навчання: виділення основних закономірностей між параметрами.

2. Етап використання: перевірка на точність залежності, що виявлена на етапі моделювання.

Щоб оцінити степінь ефективності моделі, необхідним є використання контрольної вибірки. Важливим критерієм є необхідність покриття всієї наявної інформації тестовою вибіркою.

Навчальна здатність алгоритму оцінюється шляхом порівняння точності моделювання на тестовій та контрольній вибірках. Якщо похибка на навчальній вибірці невелика і майже дорівнює значенню похибки на контрольній вибірці, то це говорить про гарну спроможність до навчання алгоритму.

В іншому разі за достатньо великого відхилення від фактичних значень, отриманих на тестовій вибірці, говорять про наявність такого процесу, як «недонавчання», а коли похибка на контрольній вибірці значно вища за похибку тестової вибірки, кажуть про «перенавчання» алгоритму. Неможливо повністю позбавитися «перенавчання», але можливо мінімізувати його ефект на вихідний результат.

Як критерій якості будемо використовувати значення середньоквадратичної похибки (MSE), середньої абсолютної похибки (MAE) та середньоквадратичного відхилення (RMSE). Модель, для якої ці показники матимуть найменше значення, буде використана в подальшому аналізі. Формули для розрахунку похибок (1), (2), (3).

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)^2, \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n \hat{x}_i - x_i, \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)^2} = \sqrt{MAE}, \quad (3)$$

де \hat{x}_i – прогнозоване значення показника;

x_i – фактичне значення показника;

n – розмір вибірки.

Таким чином, для моделювання міждержавного міграційного процесу на основі даних методів машинного навчання побудуємо моделі, проаналізуємо точність кожної з них на тестовій та контрольній вибірках та виберемо найбільш ефективну для подальшої роботи.

Країни, використані для аналізу, були відібрані за чисельністю міграційних потоків з Україною, причому досліджувались як потоки прибуття, так і потоки вибуття. Таким чином, було відібрано 14 країн: Азербайджан, Білорусь, Грузія, Італія, Китай, Литва, Німеччина, Польща, РФ, США, Туреччина, Узбекистан, Франція, Україна, Чехія.

Чинниками впливу на міграційний процес вибрано 20 факторів (рис. 1), з яких на основі кореляційно-регресійного аналізу було відібрано 14 найбільш значимих: ВВП на душу населення, індекс безробіття, рівень смертності, демографічне навантаження на душу населення, рівень зайнятості, індекс голоду, індекс споживчих цін, коефіцієнт народжуваності, відношення чоловіків до жінок, індекс людського капіталу і-ї країни, індекс бідності і-ї країни, витрати держави на охорону здоров'я, рівень злочинності, індекс тероризму.

На основі інформації Світового банку даних, офіційних служб статистики аналізованих країн та даних Євростату було зібрано вхідні дані, що являють собою сукупність багатомірних часових рядів для кожної країни. Складність полягає у тому, що для деяких показників характерним є наявність пропущених значень. Існує декілька способів позбавитися пропущених значень: видалення рядків або стовпців, заміна на середнє, моду або іншу константу (зазвичай використовується 0) або апроксимація за допомогою наявних значень у вибірці. Недоліком першого способу є можливість втрати значної частини інформації, що призводить до того, що вибірка буде нерепрезентативною. Заміна пропущених значень на константу може призвести до збільшення дисперсії та квадратичного відхилення. Тому було застосовано метод прогнозування відсутніх значень за допомогою вбудованої функції середовища *RStudio* – *tmisc*, що дає змогу побудувати модель за наявною інформацією,

а потім заповнити пропуски, генеруючи відсутні дані на основі побудованої моделі. Для заповнення пропущених значень було використано метод побудови дерев рішень *cart*. Отримана таким чином база даних не містить пропусків і може використовуватися для подальшого моделювання.

Як залежну змінну будемо використовувати значення відносного міграційного сальдо, що розраховується як частка від ділення чисельності потоку прибуття до країни до потоку вибуття, а як незалежні змінні – значення факторів впливу на міграційний процес.

Поділимо нашу вихідну базу даних на дві частини: навчальну та контрольну вибірку у пропорції 80/20, щоб оцінити якість моделей, побудованих методами машинного навчання. Спершу побудуємо модель, використовуючи навчальну вибірку, розрахуємо похибки моделі і перевіримо результати на контрольній вибірці. Сильна різниця між помилками навчальної та тестової вибірок свідчатиме про недостатню ефективність і продуктивність моделі.

Спершу побудуємо модель із використанням методу беггінгу – *Random Forrest*. «Випадковий ліс» (*Random Forrest*) – спосіб усереднення множини дерева рішень, що навчаються на різних частинах одного набору даних для подолання проблеми одного дерева рішень. Таким чином, метод являє собою навчання ансамблю дерев рішень для вирішення задачі класифікації та регресії, що полягає у побудові значної кількості дерев рішень для випадкових вибірок із вхідної бази даних.

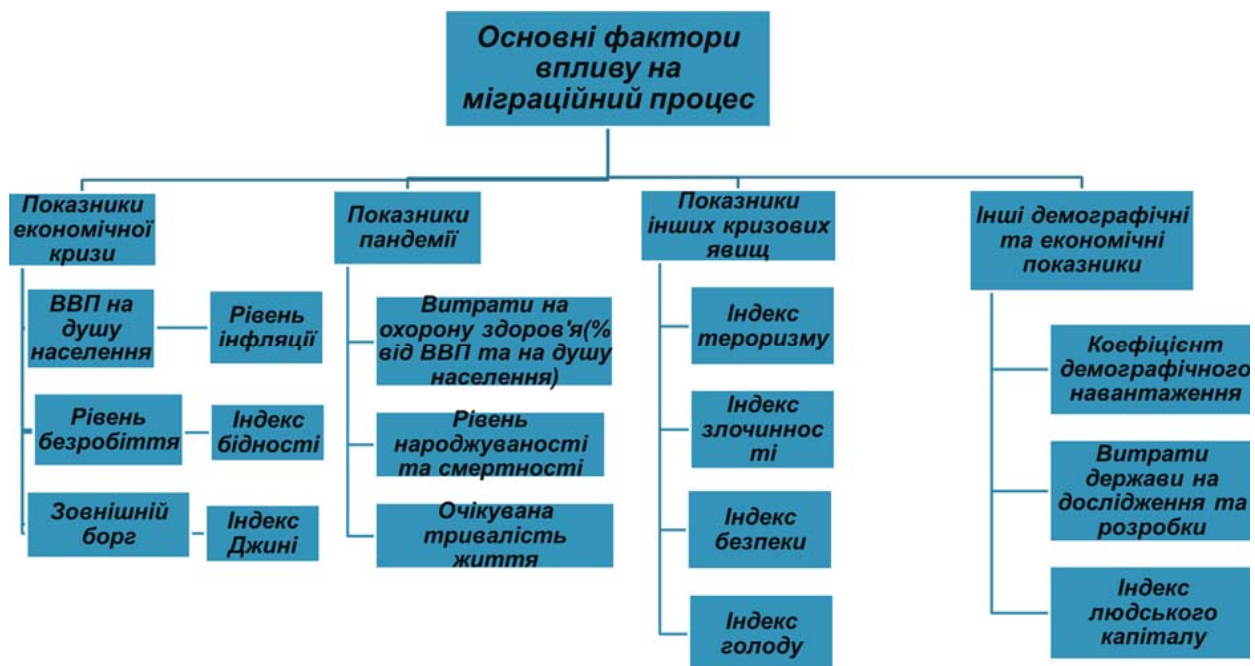


Рис. 1. Чинники впливу на зовнішній міграційний процес

Оскільки залежна змінна має кількісний вираз, ми маємо справу з задачею регресії. Для побудови моделі в *Rstudio* за допомогою алгоритму «випадкового лісу» скористаємося функцією *Randomforest* з однойменного пакету.

Таким чином, у середовищі *RStudio* будуємо модель «випадкового лісу». Припускається, що «ліс» передбачає середнє значення для відносного міграційного сальдо, орієнтуючись на значення факторів, що впливають на нього. Для моделі було протестовано різні значення вхідних параметрів для мінімізації похибки. Найбільшу точність модель продемонструвала за кількості дерев: $Ntree = 40$, кількості параметрів для розбивки: $Mtry = 8$.

Для перевірки точності моделі, спрогнозуємо значення залежної змінної для того, щоб порівняти з фактичним, та розрахуємо значення похибок обчислень. Щоб знайти значення міграційного сальдо, розрахованого за допомогою моделі *Random Forest*, скористаємося функцією *Predict*.

Побудовану на тестовій вибірці модель було перевірено на контрольній, значення помилок, отримані в результаті, були приблизно однакові, що свідчить про адекватність моделі, отже, її можна використовувати для прогнозування міграційного процесу. Похибки обчислень представлено в табл. 1.

Після побудови моделі візуалізуємо результати. Якщо проаналізувати графік порівняння вхідних даних зі значеннями, отриманими в

результаті використання алгоритму, можна побачити, що значення майже накладаються одні на одних, що свідчить про достатню точність моделі (рис. 2).

Таблиця 1

Похибки моделі Random Forest			
Похибка	MSE	MAE	RMSE
Значення похибки	0.119308	0.158284	0.361398

Для побудови другої моделі використаємо *XGBoost* – метод, що відноситься до ансамлевої моделі навчання – бустінгу. Extreme Gradient Boosting (*XGBoost*) – бібліотека з відкритим кодом, яка забезпечує ефективну та дієву реалізацію алгоритму підвищення градієнту. Ансамблі будуються з моделей дерев рішень, дерева додаються по одному і так, щоб виправити помилки прогнозу, що були зроблені попередніми моделями.

Алгоритм використовує будь-яку довільну функцію втрат та алгоритм оптимізації градієнтного спуску, завдяки чому метод і отримав свою назву – «підвищення градієнту», оскільки градієнт втрат мінімізується в процесі моделювання.

У середовищі *RStudio* цей метод реалізується за допомогою пакету *XGBoost*, основними перевагами його застосування є швидкість виконання моделі та її продуктивність.

У ході дослідження було протестовано алгоритми з різними параметрами за допомогою

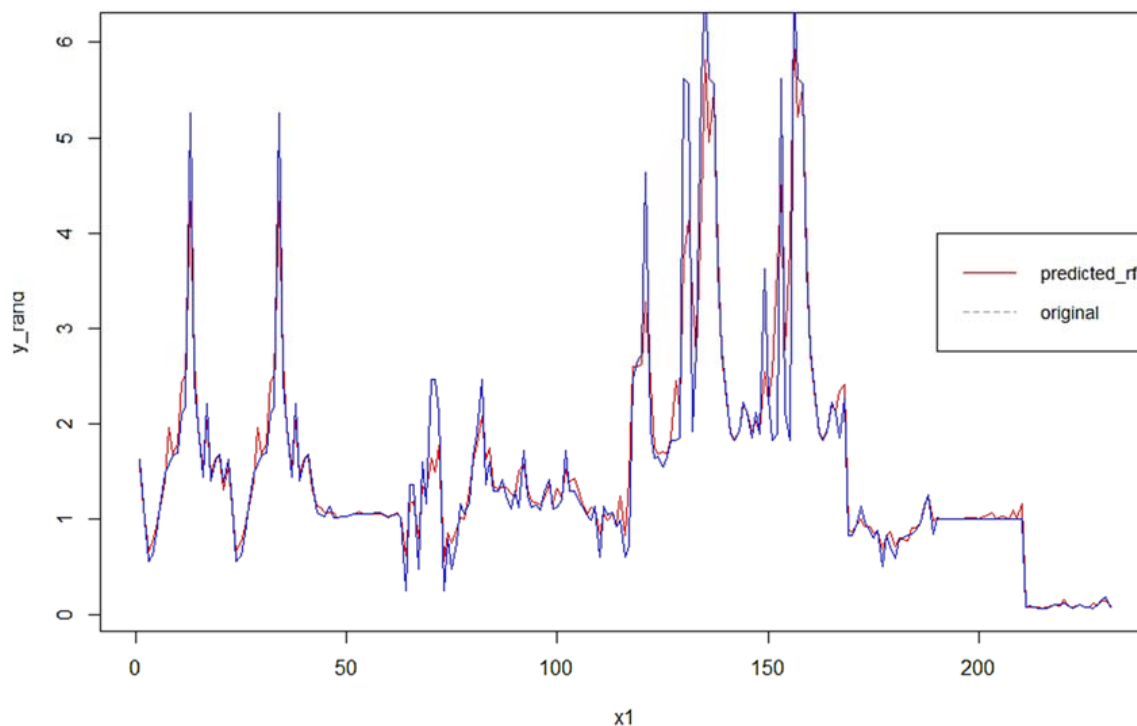


Рис. 2. Візуалізація фактичного та прогнозованого значень залежної змінної

функції *grid_tune*, що дає змогу підбирати різні атрибути для моделювання задля досягнення максимальної точності. Щоб зменшити похибку обчислень та збільшити точність прогнозування моделі, було підбрано спочатку кількість дерев, що використовуються в аналізі, потім – максимальну глибину та параметр, що контролює швидкість навчання.

У процесі виконання роботи було перевірено похибки виконання моделі на тестовій та контрольній вибірках, їхні значення були близькі між собою, отже, модель підлягає подальшому використанню. Під час аналізу похибок було виявлено, що значення середньої абсолютної та середньої квадратичної похибки достатньо невеликі (0,43 та 0,72 відповідно), тоді як значення коефіцієнта детермінації близьке до 1. Результати моделювання представлено на рис. 3.

RMSE	Rsquared	MAE
0.7170262	0.645965	0.4277893

Рис. 3. Оцінка результатів моделювання

Після побудови моделі було спрогнозовано значення залежної змінної за аналогією з методом Random Forest, використовуючи функцію *predict*, та візуалізовано порівняння фактичного значення відносного міграційного сальдо з розрахованим за допомогою моделі (рис. 4).

Перейдемо до побудови моделі за допомогою методу нейронних мереж. Нейронна мережа являє собою розрахункову нелінійну модель, що заснована на нейронній структурі мозку та може навчитися виконувати такі завдання, як класифікація, прогнозування,

прийняття рішень, візуалізація та ін. Штучна нейронна мережа складається зі штучних нейронів або елементів обробки і організована у три взаємопов'язані шари: вхід, прихований шар, який може складатися з декількох рівнів, та вихід. Таким чином, нейронна мережа дає змогу отримувати прогноз для цільової функції шляхом аналізу взаємозв'язку факторів та результуючої змінної у прихованому шарі за допомогою зважених з'єднань нейронів.

У *RStudio* для побудови моделі залежності одних змінних від інших існує пакет *neuralnet*, що дає змогу вирішити задачу регресії. Моделювання за допомогою нейронних мереж проходить в декілька етапів:

1. Нормалізація вхідних даних, тому що обов'язковою умовою для використання нейронних мереж є потрапляння даних у проміжок [0,1]. Нормалізація даних є важливим кроком, що дає змогу уникнути помилок, коли одна змінна сильно впливає на іншу через свій масштаб.

2. Поділ тестової вибірки на дві частини випадковим чином.

3. Побудова моделі нейронної мережі за допомогою функції *neuralnet*.

Для нормалізації вхідних даних знайдемо максимальне та мінімальне значення для кожного стовпця факторів та використаємо формулу нормалізації *min-max* для представлення кожного елемента (4).

$$x_{ij} = \frac{x_{ij} - \min x_j}{\max X_j - \min X_j}, \quad (4)$$

де x_{ij} – відповідний елемент матриці вхідних даних;

$\max(\min) X_j$ – максимальне (мінімальне) значення по стовпцю.

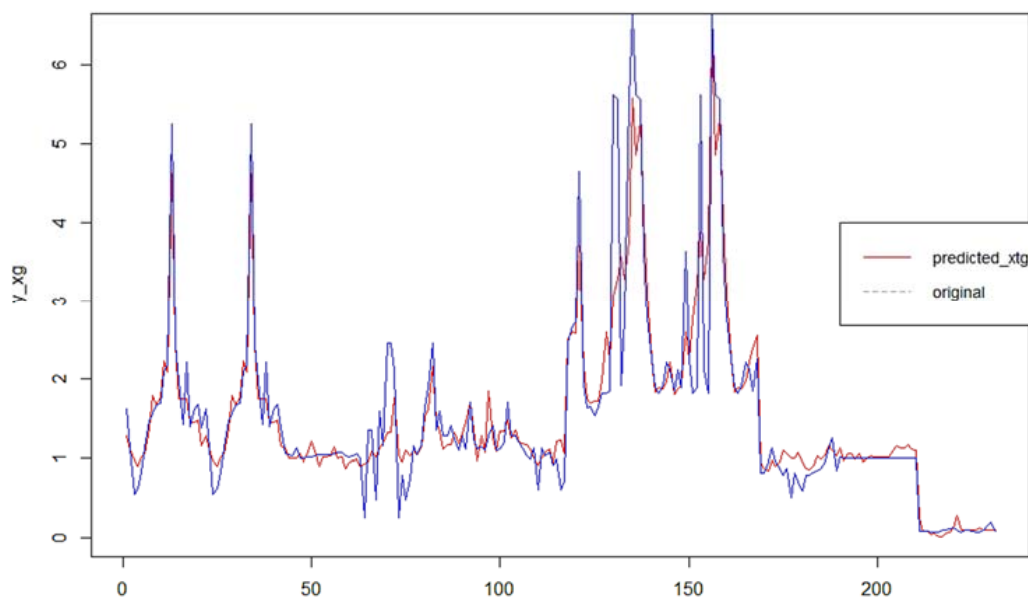


Рис. 4. Порівняння фактичних значень і змодельованих методом XGBoost

Перевагою методу нормалізації *min-max* серед усіх інших методів є те, що він дає змогу зберегти вихідний розподіл змінних.

Підберемо кількість прихованих шарів та кількість нейронів на кожному шарі так, щоб мінімізувалася похибка моделі. Таким чином, було підібрано нейронну мережу з двома прихованими шарами: два нейрони на одному шарі і п'ять нейронів на другому шарі.

Відобразимо побудовану нейронну мережу за допомогою вбудованої функції *plot*. Результат виконання відображено на рис. 5.

Для того щоб порівняти отримані значення цільової функції з фактичними, необхідно спо-

чатку повернутися до нормального вигляду. Для цього помножимо отриманий вектор на різницю між максимальним та мінімальним елементами по стовпцю та додамо мінімальний.

Під час розрахунку похибок моделі, використавши отримане обчислене значення, на тестовій та контрольній вибірках отримали достатньо малі значення, причому під час порівняння помилок від моделювання на обох вибірках не було помічено суттєвої різниці. Для візуального порівняння фактичних значень із розрахованими за допомогою нейронної мережі побудуємо графік (рис. 6).

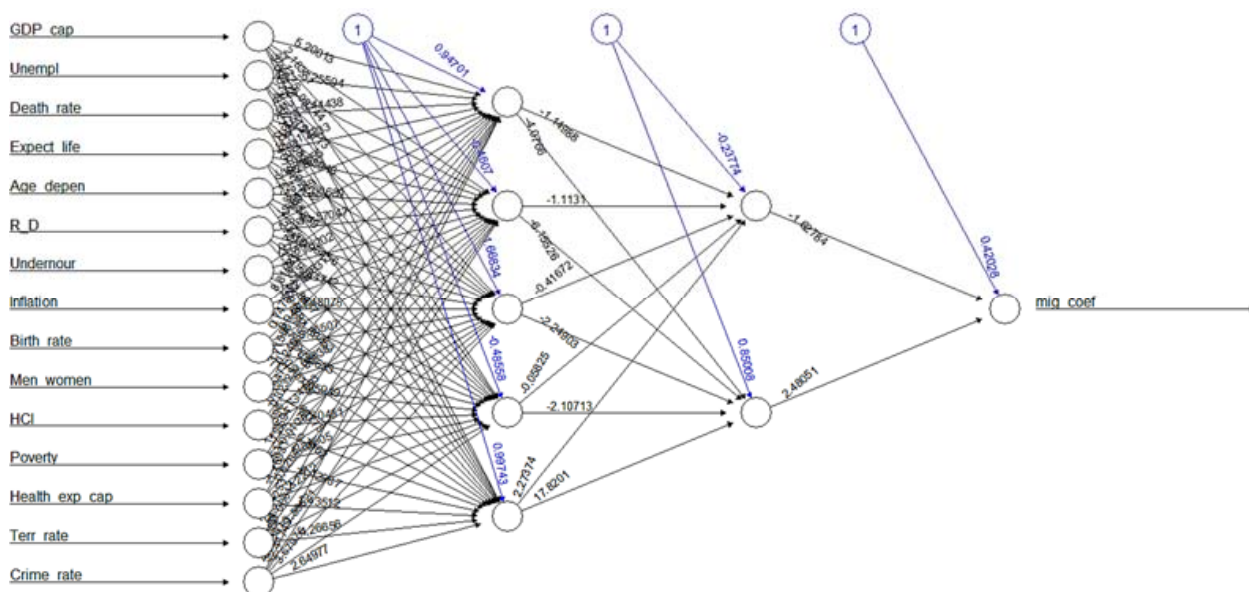


Рис. 5. Схема нейромережі для вхідних параметрів

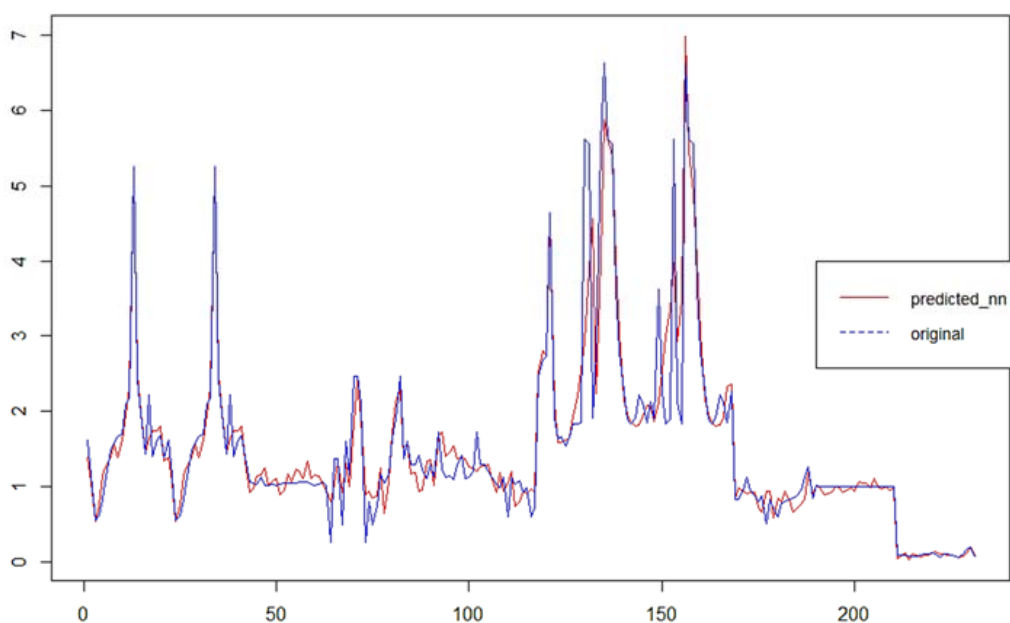


Рис. 6. Порівняння фактичного результату і змодельованого нейромережею

Порівнюємо похибки, отримані під час використання кожної з моделей, та виберемо модель для подальших досліджень (табл. 2).

Таблиця 2

**Порівняння похибок
при використанні різних моделей**

	MSE	MAE	RMSE
RF	0.119308	0.158284	0.361398
XTGboost	0.266508	0.295249	0.516245
NN	0.179966	0.213448	0.424224

Найменші значення похибок були отримані внаслідок використання методу ансамблевого навчання – *Random Forrest*, тому саме його будемо використовувати для прогнозування міграційного процесу на Україні.

Для прогнозування значень факторів впливу на міграцію скористаємося методами передбачення часових рядів. Щоб знайти наступні два значення для кожного показника моделі, використаємо модель *ARIMA (p, n, q)*.

Щоб відібрати найкращу модель, скористаємося функцією бібліотеки *Predict – auto.arima*, як інформаційний критерій підбору використаємо мінімальний критерій Акаїке. Результати передбачення наступних двох значень критеріїв відображено в табл. 3.

Перейдемо до прогнозування значення відносного міграційного сальдо на 2021–2022 рр., скориставшись методом *Random Forest*. У результаті отримуємо значення, що представлені в табл. 4.

Таблиця 3

Прогнозування значень факторів впливу

Назва фактору	2021	2022
GDP_cap	3662.56	3726.93
Unempl	8.19	9.48
Death_rate	14.70	16.00
Expect life	71.83	68.08
Age_depen	48.47	49.12
R_D	0.47	1.02
Undernour	2.50	2.50
Inflation	7.89	2.73
Birth_rate	8.10	9.00
Men_women	0.86	0.86
HCI	0.78	0.63
Poverty	0.20	2.20
Health_exp_cap	157.54	52.88
Terr_rate	5.55	4.69
Crime_rate	48.88	48.85

Таблиця 4

**Прогнозовані значення міграційного сальдо
для України**

Рік	Міграційне сальдо
2021	1.645316
2022	1.307047

Ілюстрація отриманих результатів для України відображена на рис. 7, де зображено графік із фактичними значеннями залежної змінної, значення, отримані з моделі, та прогноз на два періоди вперед.

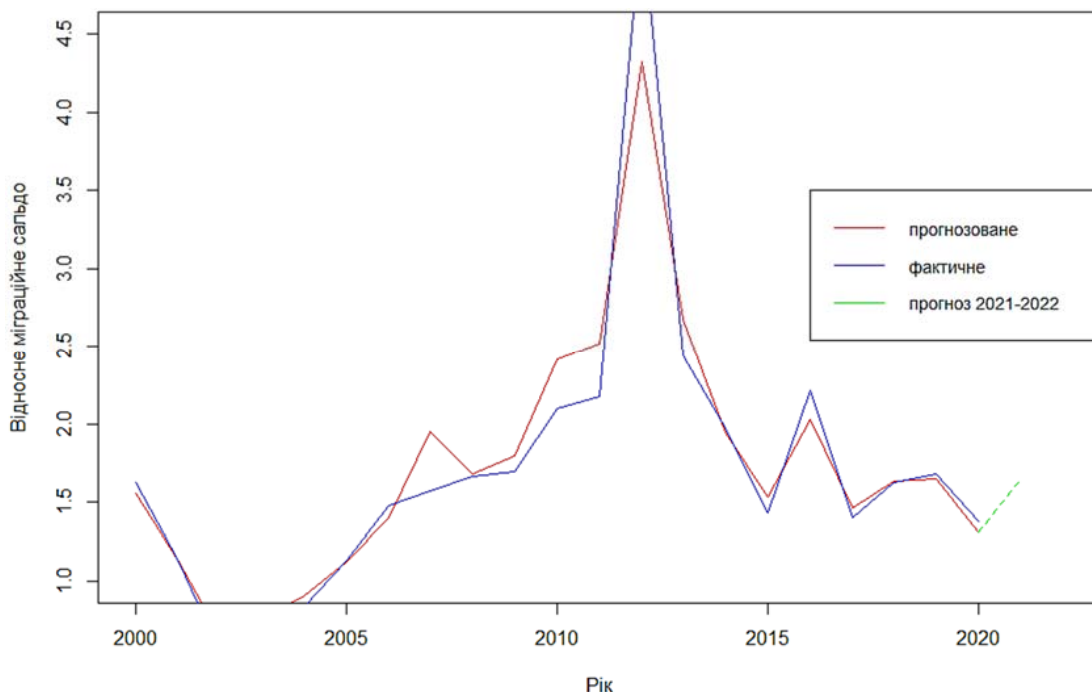


Рис. 7. Порівняння фактичного значення міграційного сальдо з прогнозом

Висновки з проведеного дослідження.

Таким чином, у статті проведено дослідження міграційного процесу, що охоплював 15 країн. Для аналізу було виділено країни, що характеризуються найбільшими міграційними потоками з Україною.

Для оцінки зовнішньої міграції кожної з країн було враховано вплив на них факторів, що є показниками економічної кризи, пандемії та інших кризових явищ. Наряду з чинниками, пов'язаними з кризовими ситуаціями, для аналізу враховувалися також важливі економічні та демографічні індикатори.

Під час моделювання міграційного процесу як залежну змінну було використано відносне міграційне сальдо, а як незалежні змінні були застосовані значення факторів впливу на міграційний процес. Таким чином, було побудовано три моделі за допомогою методів машинного навчання, які навчалися на статистичних даних

по 15 країнам, взятих за 20 років. Після аналізу похибок обчислень кожної з моделі найнижчі показники були отримані зі використання методу *Random Forest*, тому саме його було використано у подальшому аналізі.

Таким чином, було виконано прогноз значень факторів впливу на міграційний процес для України, використовуючи моделі ARIMA для прогнозування часових рядів. Отримані передбачені значення показників були використані для прогнозування міграційного процесу на 2021–2022 рр. за допомогою ансамблевого методу машинного навчання *Random Forest*. У результаті отримано прогноз відносного міграційного сальдо для України. Отриману модель можна використовувати під час прогнозування зовнішнього міграційного процесу рухів під дією соціально-економічних кризових явищ та інших економічних показників, щоб приймати важливі стратегічні рішення для держави.

Список використаних джерел:

1. Хавинсон М.Ю., Кулаков М.П. Концепция динамической гравитационной модели миграции населения. *Региональные проблемы*. 2016. № 4. С. 12–19.
2. Москвіна В.А., Ровенчак О. Моделирование межрегиональной мобильности выпускников вузов России. *Прикладная эконометрика*. 2019. № 56. С. 99–122.
3. Антосик Л.В. Факторы и направления межрегиональной миграции выпускников вузов России. *Вопросы образования*. 2012. № 2. С. 107–125.
4. Антосик Л.В., Ивашина Н.В. Моделирование пространственной зависимости миграционных потоков выпускников вузов РФ. *Прикладная эконометрика*. 2019. № 54. С. 70–89.
5. Christan Alis, Erika Fille Legara, Christopher Monterola. Generalized radiation model for human migration. *Scientific reports*. 2021. № 11.
6. Стебунова О.И. Моделирование и сценарное прогнозирование внешней трудовой миграции в условиях вступления России в ВТО. *Вестник ОГУ*. 2014. № 14(175). С. 465–470.
7. Батищева Г.А., Журавлева М.И. Математическое моделирование внешних российских миграционных потоков. *Проблемы экономики и менеджмента*. 2015. № 6(46). С. 28–34.
8. Hosseini S.I., Tarasyev A., Machine learning methods in individual migration behavior. *Business informatics and mathematical modelling in the digital economy*. P. 72–81.
9. Yosha Bengio. Learning Deep Architectures for AI / Foundations and Trends in Machine Learning, 2009.
10. Документація середовища RStudio. URL: <https://www.rdocumentation.org/>.

References:

1. Khavinson M.Yu., Kulakov M.P. (2016) Kontseptsiya dinamicheskoy gravitatsionnoy modeli migratsii naseleniya. *Regional'nyye problemy*, no. 4, pp. 12–19.
2. Moskvina V.A., Rovenchak O. (2019) Modelirovaniye mezhregional'noy mobil'nosti vypusknikov VUZov Rossii. *Prikladnaya ekonometrika*, no. 56, pp. 99–122.
3. Antosik L.V. (2012) Faktory i napravleniya mezhregional'noy migratsii vypusknikov vuzov Rossii. *Voprosy obrazovaniya*, no. 2, pp. 107–125.
4. Antosik L.V., Ivashin N.V. (2019) Modelirovaniye prostranstvennoy zavisimosti migratsionnykh potokov vypusknikov vuzov RF. *Prikladnaya ekonometrika*, no. 54, pp. 70–89.
5. Christan Alis, Erika Fille Legara, Christopher Monterola (2021) Generalized radiation model for human migration. *Scientific reports*, no. 11.
6. Stebunova Ye.I. (2014) Modelirovaniye i stsenarnoye prognozirovaniye vneshey trudovoy migratsii v usloviyakh vstupleniya Rossii v VTO. *Vestnik OGU*, no. 14(175), pp. 465–470.
7. Batishcheva G.A., Zhuravleva M.I. (2015) Matematicheskoye modelirovaniye vneshnikh russkikh migratsionnykh potokov. *Trudnosti ekonomiki i menedzhmenta*, no. 6(46), pp. 28–34.
8. Hosseini S.I., Tarasyev A. Machine learning methods in individual migration behavior. *Business informatics and mathematical modelling in the digital economy*. P. 72–81.
9. Yosha Bengio (2009) Learning Deep Architectures for AI. Foundations and Trends in Achine Learning.
10. Dokumentatsiya seredovishcha RStudio. URL: <https://www.rdocumentation.org/>.