

В.Г. Гуськова, аспірант,  
П.І. Бідюк, д-р техн. наук, проф.  
Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна  
guskovavera2009@gmail.com, pbidyuke\_00@ukr.net

## Розробка сценарного підходу на основі моделей інтелектуального аналізу даних

*Дана робота присвячена використанню статистичних показників від Організації Об'єднаних Націй та Всесвітнього банку для аналізу розвитку людини. Для побудови коротко-, середньо- та довгострокових прогнозів були розроблені такі математичні моделі, як авторегресійне рівняння, авторегресійне рівняння з ковзним середнім (де ковзне середнє побудовано по залишкам авторегресійної моделі), авторегресійна модель з ковзним середнім із власними коефіцієнтами при ковзному середньому. Альтернативними методами для побудови прогнозу були нейронна мережа з радіально-базисними функціями та нечіткий метод групового урахування аргументів. За результатами побудови моделей та прогнозу було проведено аналіз якості та обрана модель з найкращими показниками, на основі якої розроблено сценарії подальшого розвитку країн.*

**Ключові слова:** розвиток людини, прогнозування, авторегресія, ковзне середнє, нейронна мережа, радіально-базисні функції, нечіткий метод, сценарії розвитку

### Вступ

Сьогодні світ входить в епоху криз та нестачі ресурсів, тому життєво необхідним фактором є стратегічне прогнозування і планування розвитку країни, в якій спостерігається слабостійка економіка, мається непопулярна влада і нестабільне суспільство і яка може стати об'єктом для масштабного політичного експерименту. Необхідно вміти оцінювати і прогнозувати тенденції розвитку країни, а також можливі загрози для її економіки та соціального розвитку.

Поняття індексу розвитку людського потенціалу як методу вимірювання людського розвитку було введено Програмою розвитку ООН (ПРООН) в 1990 р. в Доповіді про розвиток людини [1]. На відміну від попередніх теорій, концепція людського розвитку сфокусована на людину і проголошує добробут людини основною і єдиною метою розвитку. Індекс розвитку людського потенціалу – розрахунковий статистичний показник, в якому враховуються не тільки обсяги споживання матеріальних благ, а й можливості для розвитку людини, що забезпечуються системами охорони здоров'я та освіти [1-2].

Кількість даних, пов'язаних з людським розвитком, безперервно зростає, а самі дані стають складнішими і різноманітнішими, тому важливим етапом стає їх аналіз. Методи регресійного моделювання та прогнозування даних, нейронні мережі, тощо, на сьогодні є надзвичайно важливими, адже вони цілком і повністю підходять для урахування динаміки протікання спонтанних процесів та явищ [2].

### Аналіз якості моделі та прогнозу

При побудові математичної моделі, на основі якої будуть розроблятися прогнози, проводиться аналіз на якість моделі, виконується перевірка моделей на адекватність [3]. До показників якості моделі у роботі віднесено:

- суму квадратів залишків

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2; \quad (1)$$

- коефіцієнт детермінації, який вказує наскільки отримані спостереження підтверджують модель

$$R^2 = 1 - \frac{V(y|x)}{V(y)}, \quad (2)$$

де  $V(y|x) = \sigma^2$  – умовна дисперсія змінної;

- критерій Дарбіна - Уотсона ( або DW-критерій) - статистичний критерій, який використовується для тестування автокореляції першого порядку елементів досліджуваної послідовності

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \approx 2(1 - \rho_1). \quad (3)$$

де  $\rho_1$  - коефіцієнт автокореляції першого порядку.

Важливим етапом прогнозування є верифікація прогнозів, тобто оцінювання їх точності та обґрунтованості. На етапі верифікації використовують сукупність критеріїв, способів і процедур, які дають можливість оцінити якість прогнозу [3-4]. В роботі показниками якості прогнозу обрано:

- середньоквадратичну помилку моделі:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}; \quad (4)$$

- нормовану помилку RMSE ( Root Mean Squared Error ) :

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}; \quad (5)$$

- абсолютну помилку прогнозу, яка визначена як різниця між фактичним значенням  $y_t$  і прогнозом  $y^*$ :

$$\Delta_{pr} = y_t - y^*; \quad (6)$$

- середнє абсолютне значення помилки

$$\bar{\Delta}_{pr} = \frac{\sum_{t=1}^n |y_t - y_t^*|}{n}; \quad (7)$$

- середньоквадратичну помилку прогнозу

$$\sigma_t = \sqrt{\frac{\sum_{t=1}^n (y_t - y_t^*)^2}{n}}; \quad (8)$$

- коефіцієнт нерівності Тейла – індикатор якості моделі і прогнозу. За визначенням  $0 \leq U \leq 1$ . Якщо  $U = 1$ , то модель має практично нульові (неприйнятні) прогнозуючі властивості, що випливає з формули для обчислення  $U$ :

$$U = \frac{\sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2}}{\sqrt{\frac{1}{N} \sum_{k=1}^N y^2(k)} + \sqrt{\frac{1}{N} \sum_{k=1}^N \hat{y}^2(k)}} \quad (9)$$

При  $U = 0$  прогнозовані значення співпадають з фактичними значеннями ряду.

### Прогнозування за допомогою авторегресійної моделі

Авторегресійна (AR-) модель - модель часових рядів, в якій значення часового ряду в даний момент лінійно залежить від попередніх значень цього ж ряду. Авторегресійний процес порядку  $p$  (AR( $p$ )- процес) визначається наступним чином:

$$X_t = c + \sum_{i=1}^p a_i X_{t-i} + \varepsilon_t \quad (10)$$

де  $a_i$  - параметри моделі (коефіцієнти авторегресії),

$c$  - постійна (часто для спрощення передбачається рівний нулю),

$\varepsilon_t$  - білий шум [4].

Розглянемо побудову прогнозів на прикладі Канади. Для роботи були використані дані з 1989 по 2014 рр. Прогноз будується до 2019 р. Для мінімізації похибки прогноз буде будуватися кожен раз на один крок методом зсуву вибірки. Наприкінці отримуємо вибірку з 1989 по 2019 та у подальших дослідженнях можемо використовувати дані для розбиття на навчальну та перевіірочну вибірки у співвідношенні:

- 1989 – 2014 – навчальна вибірка, 2015 – перевіірочна (прогнозування на 1 крок).

- 1989 – 2014 – навчальна вибірка, 2015 – 2017 – перевіірочна (прогнозування на 3 кроки).

- 1989 – 2014 – навчальна вибірка, 2015 – 2019 – перевіірочна (прогнозування на 5 кроків).

*simpl 1989 2014*

*kanada.correl*

*equation ar1kan.ls*

(11)

*kanada=c(1)+c(2)\*kanada(-1)+c(3)\*kanada(-2)+c(4)\*kanada(-3)*

Після побудови моделі отримуємо показники:

- R-squared = 0.7384;

- Mean development var = 0.9415;

- Akaike = -4.606;

- Darbin-Watson = 1.9919.

На основі прогнозу за допомогою авторегресійної моделі можемо використовувати цю вибірку для подальшої роботи, а саме, для прогнозування за допомогою авторегресійних моделей з ковзним середнім та нейронних мереж.

Одна з математичних моделей, яка використовується для аналізу і прогнозування стаціонарних часових рядів в роботі - авторегресійна модель з ковзним середнім:

$$X_t = c + \sum_{i=1}^p a_i X_{t-i} + \varepsilon_t + \sum_{i=1}^p \beta_i \varepsilon_{t-i} \quad (12)$$

де  $c$  - константа,  $\varepsilon_t$  - білий шум, тобто послідовність незалежних і однаково розподілених випадкових величин (як правило, нормальних).

### Прогнозування за допомогою нейронних мереж і нечітких методів

В роботі використовувався спеціальний тип нейронних мереж з прямими зв'язками, які орієнтовані на апроксимацію та інтерполяцію багатовимірних функцій для вирішення, зокрема, задач прогнозування. RBF-мережі мають тільки один шар прихованих нейронів. Єдиний вихідний нейрон видає значення функції.

$$y = L(\vec{x}) = L(x_1, \dots, x_m, \dots, x_M). \quad (13)$$

Нейрони використовують гауссівську функцію всередині себе:

$$\varphi_k(\vec{x}_n) = e^{-\frac{\|\vec{x}_n - \vec{c}_k\|^2}{2 \cdot \sigma_k^2}}, \quad (14)$$

де в якості метрики використовується

$$\|\vec{x}_n - \vec{c}_k\|^2 = \sum_{m=1}^M (x_{n,m} - c_{k,m})^2, \quad i = 1..M. \quad (15)$$

Таким чином, за своєю будовою RBF-мережі - це двошарові мережі першого порядку, причому обидва шари з'єднані ваговою матрицею  $W$ .

Для побудови часткової моделі за допомогою нечіткого методу групового урахування аргументів (НМГУА)

використовувалася лінійна інтервальна регресійна модель, що задається таким чином [4]:

$$Y = B_1 z_1 + B_2 z_2 + \dots + B_n z_n, \quad (16)$$

де  $z_i$  – деякі відомі змінні (фактори впливу),

$B_i$  – інтервали, які можна задати трикутними нечіткими числами, і записати у вигляді центра  $\alpha_i$  і ширини  $c_i$ .

Загальний опис алгоритму

1. Вибір виду моделі.
2. Вибір зовнішніх критеріїв оптимальності (критерій регулярності, незміщеності або RMSE).
3. Вибір загального виду опорної функції (виду часткових опісів).
4. Розбиття вибірки на навчальну  $N_{\text{навч}}$  і перевірку  $N_{\text{перев.}}$ .
5. Присвоєння нульових значень лічильнику кількості моделей  $k$  і лічильник числа рядів  $r$ .
6. Генерування нової часткової моделі  $f_k$  на навчальній вибірці. Розв'язання задачі лінійного програмування для шуканих  $\alpha_i$  та  $c_i$ .
7. Визначення по тестовій вибірці значення зовнішнього критерію.
8.  $k=k+1$ . Якщо  $k \geq C_F^2$ , то  $k=0$ ,  $r=r+1$ .
9. Обчислення середнього критерію для моделі  $r$ -ї ітерації. Якщо  $r=1$ , то перехід на крок 6, інакше – на крок 10.
10. Якщо  $n_{\text{см}}(r) - n_{\text{см}}(r-1) \leq \varepsilon$ , то перехід на крок 11, інакше – відбираємо  $F$  кращих моделей, і припускаючи, що  $r=r+1$ ,  $k=1$ , перехід на крок 6 і виконання наступної  $(r+1)$ -ї ітерації.
11. Вибір із  $F$  моделей попереднього ряду найкращої моделі по критерію регуляризації.

### Експериментальні дослідження

Для порівняння результатів прогнозування за допомогою регресійних моделей, нечіткого методу групового урахування аргументів та нейронної мережі радіально-базисних функцій необхідно скласти таблицю показників якості моделей та прогнозу (табл. 1).

Таблиця 1. Порівняльний аналіз показників моделей на 1 крок за усіма методами

|   | R2     | SSR    | DW    |
|---|--------|--------|-------|
| АРКС(1,1), із застосуванням власного простого КС, при N=5.        | 0.4042 | 0.0052 | 1.911 |
| АРКС(1,1), із застосуванням власного експоненційного КС, при N=5. | 0.126  | 0.0077 | 1.906 |
| АРКС(1,1), із застосуванням власного простого КС по у, при N=5.   | 0.376  | 0.0055 | 1.812 |
| РБФ 5 нейронів  | 0.1884 | 0.0028 | 0.179 |
| РБФ 10 нейронів   | 0.761  | 0.003  | 2.224 |

На основі отриманих моделей було побудовано прогнози на 1 крок. У табл. 2 наведені результати якості прогнозу при прогнозуванні на 1 крок.

При прогнозуванні на 1 крок найкращі результати було отримано нечітким методом

групового урахування аргументів. Всі значення показників є мінімальними. Показники якості моделі для прогнозування на 3 кроки наведені у табл.3.

Таблиця 2. Порівняльний аналіз показників прогнозу на 1 крок

|   | RMSE   | MAPE (%) | Theil  |
|---|--------|----------|--------|
| АРКС(1,1), із застосуванням власного експоненційного КС | 0.0020 | 0.2203   | 0.0011 |
| РБФ 5 нейронів  | 0.0285 | 3.1491   | 0.0155 |
| РБФ 10 нейронів   | 0.0082 | 0.9021   | 0.0044 |
| НМГУА   | 0.0007 | 0.0823   | 0.0004 |

Таблиця 3. Порівняльний аналіз показників моделей на 3 кроки за усіма методами

| Модель  | R2    | SSR   | DW     |
|---|-------|-------|--------|
| АРКС(1,1), із застосуванням простого КС, при N=5.                 | 0.403 | 0.005 | 1.9160 |
| АРКС(1,1), із застосуванням власного експоненційного КС, при N=5. | 0.126 | 0.007 | 1.9101 |
| АРКС(1,1), із застосуванням власного простого КС по у             | 0.374 | 0.005 | 1.7993 |
| РБФ 5 нейронів  | 0.380 | 0.002 | 1.885  |
| РБФ 10 нейронів   | 0.358 | 0.002 | 1.6653 |

Таблиця 4. Порівняльний аналіз показників прогнозу на 3 кроки

|   | RMSE   | MAPE (%) | Theil  |
|---|--------|----------|--------|
| АРКС(1,1), із застосуванням власного експоненційного КС, при N=5. | 0.0018 | 0.1472   | 0.0009 |
| РБФ 5 нейронів  | 0.0345 | 3.4647   | 0.0182 |
| РБФ 10 нейронів   | 0.0015 | 1.668    | 0.0084 |
| НМГУА   | 0.0013 | 0.1417   | 0.0007 |

При побудові прогнозу на 3 кроки найкращий результат було отримано за допомогою нечіткого методу групового урахування аргументів та регресійної моделі із застосуванням власного експоненційного КС, при N=5. НМГУА показав найгіршими усі показники. У табл. 5-6 наведені показники якості моделі при прогнозуванні на 5 кроків.

Таблиця 5. Порівняльний аналіз показників моделей на 5 кроків за усіма методами

| Модель  | R2     | SSR    | DW     |
|---|--------|--------|--------|
| АРКС(1,1), із застосуванням простого КС, при N=5.                         | 0.402  | 0.005  | 1.915  |
| АРКС(1,1), із застосуванням власного експоненційного КС, при N=5.         | 0.120  | 0.007  | 1.90   |
| АРКС(1,1), із застосуванням власного простого КС по у, при N=5. Sign_pks5 | 0.3734 | 0.0055 | 1.7985 |
| РБФ 5 нейронів  | 0.076  | 0.007  | 1.454  |
| РБФ 10 нейронів   | 0.295  | 0.009  | 1.76   |

Таблиця 6. Порівняльний аналіз показників прогнозу на 5 кроків

|   | RMSE   | MAPE (%) | Theil  |
|---|--------|----------|--------|
| АРКС(1,1), із застосуванням власного простого КС по у, при N=5. | 0.002  | 0.227    | 0.0013 |
| РБФ 5 нейронів  | 0.0147 | 1.94     | 0.0078 |
| РБФ 10 нейронів   | 0.0121 | 1.03     | 0.0065 |
| НМГУА   | 0.0064 | 0.976    | 0.0034 |

При прогнозуванні на 5 кроків найкраще значення прогнозу було отримано за допомогою регресійної моделі із застосуванням власного простого КС по у, при N=5. Найгіршим був результат, отриманий при прогнозуванні моделлю РБФ 10 нейронів.

### Побудова сценаріїв

Незалежно від типу, до якого належить сценарій, загальну методику побудови можна описати за допомогою ітеративної послідовності базових кроків, що включає в себе якісне описання процесу та побудову кількісної моделі [2].

Постановку задачі сценарного аналізу людського розвитку можна описати таким чином:

- визначити базові фактори, що впливають на розвиток людини;
- встановити залежність та побудувати модель поведінки цих факторів;
- оцінити результати реалізації того чи іншого сценарію розвитку подій.

Сценарій в даному випадку – це можливий набір майбутніх подій, що визначає розвиток, динаміку, платоспроможність та надійність, а також інші фактори, які впливають на результат людської діяльності. Необхідно:

1. Використовуючи аналіз чутливості, визначити ключові фактори.

2. Розглянути можливі сценарії та комбінації ситуацій, зумовлені коливаннями цих факторів.

3. Методом експертних оцінок визначити ймовірність кожної події, що входить в сценарії.

4. За кожним сценарієм обрахувати величини можливих втрат в разі його реалізації.

Вхідні дані моделі – складові індексу людського розвитку (ІЛР) та фактори, які можуть впливати на розвиток людини в цілому. При аналізі цих стратегій слід пам'ятати, що першочерговою задачею є підвищення рівня людського життя, а вже після – супутніх факторів. Таким чином можна виділити 4 різні підходи до впровадження процедури підвищення ІЛР:

- поступове збільшення рівня людського розвитку в межах однієї країни;
  - відсутність вираженої динаміки зростання рівня людського розвитку при наявності планів такого зростання (наприклад, з урахуванням зменшення рівня безробіття);
  - зростання рівня людського розвитку за рахунок розвитку у інших країнах ;
  - відсутність моніторингу ситуації.
- Приклади сценаріїв на різних етапах такі:

Етап 1:

- збільшення робочих місць;
- підвищення рівня освіти, починаючи з середнього;
- зниження рівня CO<sub>2</sub>;
- покращення медичного обслуговування, закупка обладнання, стажування лікарів;
- збільшення термінів виплати кредитів.

Результат: Поступове вирішення проблеми.

Етап 2:

- розширення місць до фіксованої кількості;
- підтримка рівня освіти, починаючи з середнього;
- зниження рівня CO<sub>2</sub>;
- покращення рівня медичного обслуговування;
- фіксований термін для виплати кредитів.

Результат: Стабілізація без покращень.

Етап 3:

- зупинка розширення робочих місць;
- скорочення рівня стажувань для студентів та школярів;
- фіксований рівень CO<sub>2</sub> (рівень CO<sub>2</sub> зменшиться на незначну величину, а затрати будуть дуже значними);
- зупинка закупівлі медичного обладнання, але підвищення рівня медицині;
- зупинка розширення терміну для виплати кредитів

Результат: Стабілізація без покращень.

Етап 4:

- скорочення робочих місць;
- повна зупинка стажування для студентів та школярів;
- повна зупинка зменшення рівня CO<sub>2</sub>;
- спрямування медицини на вирішення можливих загроз людства (можливість з'явлення інфекційних захворювань, можливість епідемій);
- скорочення термінів для виплати кредитів.

Результат: Подальше погіршення ситуації.

Ставиться задача максимізувати значення ІЛР із врахуванням таких факторів, як соціальна захищеність, тривалість життя, рівень освіти, значення CO<sub>2</sub>, економічна ситуація в країні та рівень безробіття (табл. 7).

Таблиця 7. – Керуючі змінні ІЛР

| Змінна | Зміст                           |
|--------|---------------------------------|
| $X_1$  | Рівень безробіття               |
| $X_2$  | Рівень освіти                   |
| $X_3$  | Рівень CO <sub>2</sub>          |
| $X_4$  | Рівень медичного обслуговування |
| $X_5$  | Терміни виплати кредитів        |

Ця задача аналізується з точки зору прийняття рішень [5-6]:

1. Як досягти max рівня ІЛР при min витратах.

2. Який рівень витрат необхідно вкласти в кожен галузь, щоб досягти max покращення ситуації.

Складається значення впливу уряду на ситуацію, яка розвивається в країні. Вплив уряду:

- високий = 0,97, ~ (0,86 – 1) ;
- середній = 0,75, ~ (0,65 – 0,85);
- низький = 0,50 ~ (0,35 – 0,64);
- дуже низький = 0,2 ~ (0 – 0,34);

В якості факторів, які можуть впливати на значення ІЛР, є рівень працевлаштування ( $R_1$ ), рівень освіти ( $R_2$ ), рівень CO<sub>2</sub> ( $R_3$ ), рівень медичного обслуговування ( $R_4$ ), термін видачі кредитів ( $R_5$ ).

Складемо матрицю значень для подальшої побудови сценаріїв (табл. 8-11) для кожного із сценаріїв.

Таблиця 8. Матриця значень для побудови сценаріїв з високим рівнем впливу держави

|       | Hight | Middle | Low  | Very low |
|-------|-------|--------|------|----------|
| $R_1$ | 0.95  | 0.8    | 0.5  | 0.3      |
| $R_2$ | 0.9   | 0.7    | 0.6  | 0.45     |
| $R_3$ | 0.91  | 0.9    | 0.82 | 0.75     |
| $R_4$ | 0.97  | 0.8    | 0.7  | 0.5      |
| $R_5$ | 0.87  | 0.7    | 0.5  | 0.1      |

Таблиця 9. Матриця значень для побудови сценаріїв з середнім рівнем впливу держави

|       | Hight | Middle | Low  | Very low |
|-------|-------|--------|------|----------|
| $R_1$ | 0.85  | 0.9    | 0.7  | 0.30     |
| $R_2$ | 0.9   | 0.95   | 0.8  | 0.50     |
| $R_3$ | 0.91  | 0.9    | 0.82 | 0.75     |
| $R_4$ | 0.85  | 0.89   | 0.75 | 0.60     |
| $R_5$ | 0.85  | 0.92   | 0.7  | 0.40     |

Таблиця 10. Матриця значень для побудови сценаріїв з низким рівнем впливу держави

|       | Hight | Middle | Low  | Very low |
|-------|-------|--------|------|----------|
| $R_1$ | 0.5   | 0.7    | 0.9  | 0.8      |
| $R_2$ | 0.3   | 0.6    | 0.97 | 0.85     |
| $R_3$ | 0.91  | 0.9    | 0.82 | 0.75     |
| $R_4$ | 0.3   | 0.5    | 0.95 | 0.90     |
| $R_5$ | 0.5   | 0.6    | 0.9  | 0.8      |

Таблиця 11. Матриця значень для побудови сценаріїв з низким рівнем впливу держави

|       | Hight | Middle | Low  | Very low |
|-------|-------|--------|------|----------|
| $R_1$ | 0.5   | 0.7    | 0.8  | 0.95     |
| $R_2$ | 0.2   | 0.4    | 0.6  | 0.9      |
| $R_3$ | 0.91  | 0.9    | 0.82 | 0.75     |
| $R_4$ | 0.6   | 0.7    | 0.8  | 0.92     |
| $R_5$ | 0.65  | 0.75   | 0.89 | 0.95     |

Після складання матриці ми генеруємо сценарії. Напишемо рівняння за яким буде розраховуватися ймовірність настання сценарію при дотриманні усіх значень факторів.

При  $R_1 = H, \dots R_6 = H$ .

$$\rho(\text{hight} = H)R_1 = \frac{p(R_1 = H, R_2 = H, R_3 = H, R_4 = H, R_5 = H, R_6 = H)}{\sum \text{sit} (p (R_1 = H, R_2 = H, R_3 = H, R_4 = H, R_5 = H, R_6 = H))} \quad (17)$$

При  $R_1 = M, \dots R_6 = M$ .

$$\rho(\text{middle} = M)R_1 = \frac{p(R_1 = M, R_2 = M, R_3 = M, R_4 = M, R_5 = M, R_6 = M)}{\sum \text{sit} (p (R_1 = M, R_2 = M, R_3 = M, R_4 = M, R_5 = M, R_6 = M))} \quad (18)$$

При  $R_1 = L, \dots R_6 = L$ .

$$\rho(\text{low} = L)R_1 = \frac{p(R_1 = L, R_2 = L, R_3 = L, R_4 = L, R_5 = L, R_6 = L)}{\sum \text{sit} (p (R_1 = L, R_2 = L, R_3 = L, R_4 = L, R_5 = L, R_6 = L))} \quad (19)$$

$$\rho(VL)R_1 = \frac{p(R_1 = VL, R_2 = VL, R_3 = VL, R_4 = VL, R_5 = VL, R_6 = VL)}{\sum \text{sit} (p (R_1 = VL, R_2 = VL, R_3 = VL, R_4 = VL, R_5 = VL, R_6 = VL))} \quad (20)$$

Ймовірність першого сценарію.  $P(h) = 0.63632$ ,  $P(m) = 0.27354$ ,  $P(l) = 0.083517$ ,  $P(vl) = 0.005415$ .

$P(s1) = P(h) + P(m) + P(l) + P(vl) = 0.63632 + 0.27354 + 0.083517 + 0.005415 = 0.998792$ ,  $P(s1/h) = \frac{0.63632}{0.998792} = 0.6371896$ ,  $P(s1/m) = \frac{0.27354}{0.998792} = 0.2738$ ,  $P(s1/l) = \frac{0.083517}{0.998792} = 0.08361$ ,  $P(s1/vl) = \frac{0.005415}{0.998792} = 0.00542$ .

При дотриманні всіх пунктів із сценарію 1, ймовірність настання цього сценарію 63,71%; ймовірність настання сценарію 2 – 27,38%; сценарію 3 – 8,361% та четвертого – 0,542%. Результат було отримано адекватний, бо при

дотриманні усіх пунктів з оптимістичного сценарію ймовірність настання найгіршого має бути мінімальною [7].

Ймовірність другого сценарію  $P(h) = 0.377175$ ;  $P(m) = 0.42074$ ;  $P(l) = 0.18075$ ;  $P(vl) = 0.02025$ ;

$P(s2) = P(h) + P(m) + P(l) + P(vk) = 0.998915$ .  $P(s2/h) = 0.377849$ ;  $P(s2/m) = 0.42158$ ;  $P(s3/l) = 0.180946$ ;  $P(s4/vl) = 0.02027$ .

Перевірка:  $P(s2/h) + P(s2/m) + P(s2/l) + P(s2/vl) = 0.377849 + 0.42158 + 0.180946 + 0.02027 = 1$ .

При дотриманні всіх пунктів із сценарію 2, ймовірність його настання – 42,15%; ймовірність настання сценарію 1 – 37,78%. 3 – 18,09% та останнього 4 – 2,027%.

Аналогічним чином розраховуємо значення ймовірностей настання третього сценарію та 1, 2 та 4.  $P(h) = 0.0102$ ;  $P(m) = 0.0567$ ;  $P(l) = 0.305$ ;  $P(vl) = 0.1836$ ;  $P(s3) = P(h) + P(m) + P(l) + P(vk) = 0.5555$   $P(s3/h) = 0.01837$ ;  $P(s3/m) = 0.10216$ ;  $P(s3/l) = 0.54954$ ;  $P(s3/vl) = 0.33081$ .

Перевірка:  $P(s2/h) + P(s2/m) + P(s2/l) + P(s2/vl) = 0.01837 + 0.10216 + 0.54954 + 0.33081 = 1$ .

При дотриманні пунктів із сценарію 3, параметри якого описані вище, ймовірність його настання – 54,95%. Першого – 1,837%. Другого – 10,21%. Та останнього – 33,08%.

Ймовірність настання четвертого (песимістичного сценарію). Для цього прорахуємо ймовірності та суму настання значення. Після чого, засовувавши формулу ймовірності Байєса, отримаємо ймовірність настання кожного із сценаріїв при урахуванні того, що виконуються всі пункти із сценарію 4.

$P(h) = 0.000708$ ;  $P(m) = 0.02646$ ;  $P(l) = 0.056048$ ;  $P(vl) = 0.11208$ .

$P(s4) = P(h) + P(m) + P(l) + P(vk) = 0.19529$ ;  $P(s4/h) = 0.003625$ ;  $P(s4/m) = 0.13549$ ;  $P(s4/l) = 0.28699$ ;  $P(s4/vl) = 0.57391$ .

Перевірка:  $P(s2/h) + P(s2/m) + P(s2/l) + P(s2/vl) = 0.003625 + 0.13549 + 0.28699 + 0.57391 = 1$ .

При дотриманні песимістичного сценарію ймовірність його настання – 57,39%. Першого – 0,3%. Другого – 13,54% та третього – 28,699%. На основі розрахунків можна сказати:

- при дотриманні усіх пунктів із сценарію 1, а саме впровадження урядом високого рівня ІЛР, зменшення рівня безробіття, підвищення освіти, зменшення рівня CO<sub>2</sub>, покращення медичних послуг та обслуговування, розширення робочих посад, збільшення терміну виплат кредитів значення індексу людського розвитку буде досягати з

ймовірністю 63,718% відмітки = 0.937440, 0.938552, 0.940551, 0.937585, 0.941787.

- при дотриманні другого сценарію, а саме коли уряд буде впроваджувати високий рівень ІЛР, буде готуватися до завчасного погашення кредитів, буде підготовлювати робочі місця для розширення, рівень CO<sub>2</sub> буде продовжувати падати, буде розглянута можливість закупки нового медичного обладнання та покращення рівня освіти, рівень людського розвитку буде досягати відмітки = 0.92239, 0.91951, 0.93592, 0.92289, 0.93700 з ймовірністю 42.15%

- якщо дотримуватися сценарію III, де уряд не впроваджує високий рівень ІЛР, коли йде підготовка до продажу акцій та активів, зниження рівня CO<sub>2</sub> зупиняється, зупиняється також розширення робочих місць та зупиняється закупівля медичного обладнання, але рівень медичного обслуговування підвищується, значення рівня людського розвитку складає = 0.968287, 0.962959, 0.960203, 0.944479, 0.958628 з ймовірністю 54.95%;

- останнім етапом буде відмова урядом впроваджувати високий рівень ІЛР, зупиняється продаж частини активів, повністю зупиняється зменшення рівня CO<sub>2</sub>, зупиняються стажування для студентів та школярів, медичний сектор направляє на вирішення можливих загроз (можливість з'явлення інфекційних захворювань, епідемій), значення ІЛР буде найнижчим з ймовірністю 57.39% і дорівнювати = 0.942878, 0.942420, 0.944809, 0.943311, 0.945236;

## Висновки

Дана робота присвячена аналізу, моделюванню та прогнозуванню індексу людського розвитку за допомогою регресійних математичних моделей – АР та АРМА, нейронних мереж із радіально-базисними функціями та нечітким методом групового урахування аргументів та побудові сценаріїв подальшого розвитку. В роботі виконаний огляд розвитку людини як складової сталого розвитку. Розглянуто та проаналізовано концепцію людського розвитку та складові індексу. На основі отриманих даних була розглянута можливість розробки сценаріїв для запобігання катастроф великого масштабу, настання яких може призвести до незворотних наслідків.

## Список літератури

1. Рейтинг країн, більш всього постраждалих від кризи. За матеріалами The World's Hardest-Hit Economies [Електронний ресурс]. – Режим доступу: [http://hdr.undp.org/sites/default/files/hdr15\\_standalone\\_overview\\_ru.pdf](http://hdr.undp.org/sites/default/files/hdr15_standalone_overview_ru.pdf)

2. Саричев В. І. Індекс людського розвитку: проблеми вимірювання на національному та світовому рівнях / В. І. Саричев, Г. Ю. Єлісеєва. – К.: БІЗНЕСІНФОРМ. – 2014. – № 2. – С.220-227.
3. Бідюк П.І. Аналіз часових рядів / П. І. Бідюк// Навч. посібник. – К. : НК ПСА НТУУ КП, 2004. – 148 с.
4. Зайченко Ю.П. Нечіткі моделі і методи в інтелектуальних системах/ Ю. П. Зайченко// Навч. посібник. – К.: Слово, 2008. – 340 с.
5. Дмитриева О.А. Параллельное моделирование жестких систем на основе диагонализации полной матрицы / О.А. Дмитриева// Искусственный интеллект. – 2011. – № 4. – С. 46-53.
6. Zahirska I. Bayesian network application possibilities for corporate state modelling / I. Zahirska // Information systems, mechanics and control. - NTUU "KPI". – Vol. 4. Special issue. – Kyiv, 2010. – P.75-81
7. Згуровський М.З. Системна методика побудови БМ / М.З. Згуровський, П.І. Бідюк, О.М. Терентьєв // Наукові вісті НТУУ «КПІ». – 2007. –№4. – С. 47-61.

*Надійшла до редакції 10.11.2016*

### **В.Г. ГУСЬКОВА, П.І. БИДЮК.**

Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», Киев, Украина

### **РАЗРАБОТКА СЦЕНАРНОГО ПОДХОДА НА ОСНОВЕ МОДЕЛЕЙ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ**

Данная работа посвящена анализу развития человека с использованием статистических показателей от Организации Объединенных Наций и Всемирного банка. Для построения коротко-, средне- и долгосрочных прогнозов были разработаны такие математические модели как: авторегрессионные уравнения, авторегрессионные уравнения со скользящим средним (где скользящее среднее построено по остаткам авторегрессии), авторегрессионная модель со скользящим средним с собственными коэффициентами при скользящем среднем. Альтернативными методами для построения прогноза является нейронная сеть с радиально-базисными функциями и нечеткий метод группового учета аргументов. По результатам построения моделей и прогноза был проведен анализ качества и выбрана модель с лучшими показателями, на основе которой разработаны сценарии дальнейшего развития стран.

**Ключевые слова:** *развитие человека, прогнозирование, авторегрессия, скользящее среднее, нейронная сеть, радиально-базисные функции, нечеткий метод, сценарии развития*

### **V.H. HUSKOVA, P.I. BIDYUK**

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

### **DEVELOPMENT OF SCENARIO APPROACH BASED ON THE MODEL OF DATA MINING**

In this paper, a human development analysis using statistical indicators from international organizations, the UN and the World Bank is provided. It is based on the construction of mathematical models to analyze the data and a short, average and long-term forecasting of components of human development indicators is made. Based on statistics we built many models (AR model; ARMA, where MA is built on the remains of the AR (p) equation; ARMA with their own coefficients of the OC; ARMA to calculate coefficients, neural networks and RBF, FGMDH). High quality estimates of short and medium-term forecasts for the index of human development were obtained, and on their basis the scenario of further development of the countries was made. The examples of the methods and the analysis of the results are provided.

**Keywords:** *human development, forecasting, autoregressive, moving average, neural network, radial basis functions, fuzzy method, scenarios*