

**НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ І ГЛОБАЛЬНОГО
ІНФОРМАЦІЙНОГО ПРОСТОРУ**

КОВАЛЬ РОМАН ГРИГОРОВИЧ

004.67:004.89:[351.84:364.3-64](043.3)

**ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ЗБОРУ ТА ОБРОБЛЕННЯ ДАНИХ ДЛЯ
ПРОГНОЗУВАННЯ ВИТРАТ НА ПЕНСІЇ І СОЦІАЛЬНІ ДОПОМОГИ**

05.13.06 – Інформаційні технології

РЕФЕРАТ

дисертації на здобуття наукового ступеня
кандидата технічних наук

Київ — 2025

Дисертацією є рукопис.

Робота виконана в Інституті телекомунікацій і глобального інформаційного простору НАН України

Науковий керівник:

доктор технічних наук, професор,
член-кореспондент НАН України
ТРОФИМЧУК Олександр Миколайович,
Інститут телекомунікацій і глобального
інформаційного простору НАН України,
директор

Офіційні опоненти:

доктор фізико-математичних наук, старший
науковий співробітник,
член-кореспондент НАН України
ПЕПЕЛЯЄВ Володимир Анатолійович,
Інститут кібернетики імені В. М. Глушкова,
завідувач відділу методів системного
моделювання

доктор філософії з комп'ютерних наук,
ГУСЬКОВА Віра Геннадіївна,
Інститут прикладного системного аналізу
національного технічного університету України
«Київський політехнічний інститут імені Ігоря
Сікорського»,
доцент кафедри штучного інтелекту

Захист дисертації відбудеться 1 жовтня 2025 року о 13⁰⁰ на засіданні спеціалізованої вченої ради Д 26.255.01 Інституту телекомунікацій і глобального інформаційного простору НАН України за адресою: 03186, м. Київ, Чоколівський бульвар, 13, к. 601.

З дисертацією можна ознайомитись у бібліотеці Інституту телекомунікацій і глобального інформаційного простору НАН України за адресою: 03186, м. Київ, Чоколівський бульвар, 13.

Автореферат розісланий 29 серпня 2025 року.

Вчений секретар
спеціалізованої вченої ради
Д. Т. Н., доцент

О. М. Терентьєв

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми дослідження. Інформаційно-аналітичні системи, комп'ютерні системи підтримки прийняття рішень, системи електронного документообігу, веб-портали, тощо, активно використовуються у державному управлінні та публічному урядуванні як за кордоном, так і в Україні. Відмінністю процесів інформатизації у соціальній сфері є формування єдиного інформаційного простору — Єдиної інформаційної системи соціальної сфери України. Замість розрізнених реєстрів та баз даних впроваджується цілісна інформаційна система з єдиною програмно-технічною архітектурою, здатна ефективно адаптуватися до нових задач та викликів, що постають перед країною.

Сучасні виклики роблять особливо актуальними задачі аналізу та прогнозування потреб Пенсійного фонду України та бюджетів різних рівнів у фінансуванні витрат на виплату пенсій та соціальних допомог. Потрібні нові науково обґрунтовані підходи до розроблення інформаційних технологій, адаптованих до умов соціальної сфери, які дозволяти б виявляти потреби різних груп населення у соціальному захисті та соціальному забезпеченні, враховувати вплив чинників, що найбільше впливають на ситуацію у соціальній сфері, розглядати різні сценарії на довго-, середньо- та короткострокову перспективу, тощо. Тому, тема роботи є актуальною, має наукове та практичне значення.

Інформаційно-аналітичному забезпеченню соціальної сфери присвячені роботи як закордонних, так і вітчизняних науковців. У них розглянуто різні аспекти розроблення та впровадження інформаційно-аналітичних систем. Використанню математичних моделей та інструментів інтелектуального аналізу даних у системах підтримки прийняття рішень значну увагу приділено в роботах П. І. Бідюка, Л. Ф. Гуляницького, С. О. Довгого, Т. В. Запорожець, О. М. Трофимчука. Досвід розроблення інформаційних систем для соціальної сфери описано в роботах вітчизняних та закордонних науковців: П. Антарі, А. М. Д'аркангеліс, О. Додон, О. Коваленко, Ф. Л. Колавола, О. А. Одуволя, А. Пертиві, Ф. Р. Прадана, Ф. Ріфлі, Д. Томара, Г. П. Ситника, А. Ф. Фестаса. Забезпечення надійності та ефективності функціонування комп'ютерних систем досліджено у роботах В. А. Пепеляєва, О. М. Хіміча, М. Т. Фісуна, оброблення великих даних — у роботах М. П. Комара, Д. В. Ланде, К. Лінча.

Незважаючи на значну кількість досліджень, проблема розроблення моделей, методів та інформаційних технологій, призначених для використання у системі соціального захисту та соціального забезпечення залишається невирішеною повністю. Зокрема, слід відзначити відсутність загальних підходів до збору, аналізу та попередньої обробки структурованих і неструктурованих даних, які описують процеси, що мають місце у соціальній сфері. Нерозв'язаною залишається й низка задач розроблення універсальних методик побудови прогнозів соціально-економічних процесів, зокрема в умовах невизначеності.

Тому, важливою і *актуальною науково-прикладною проблемою*, що розв'язується в дисертаційній роботі є підвищення ефективності підтримки прийняття рішень у системі соціального захисту та соціального забезпечення в умовах невизначеності, під впливом внутрішніх та зовнішніх чинників.

Зв'язок роботи з науковими програмами, планами, темами

Тема дисертаційної роботи повністю відповідає тематиці наукових досліджень Інституту телекомунікацій і глобального інформаційного простору НАН України. Результати, отримані в ході виконання дисертаційного дослідження знайшли відображення у науково-дослідних роботах: «Розробка та аналіз засобів теоретико-ігрового моделювання стратегій збалансованого технологічного розвитку територій» (номер державної реєстрації 0116U000796), «Розробка інформаційної технології моделювання і прогнозування розвитку соціально-еколого-економічних систем в умовах невизначеності, нестаціонарності та ризику» (номер державної реєстрації 0121U100132), «Розробка інформаційних технологій та інструментальних засобів моделювання і прогнозування розвитку територій в умовах децентралізації» (номер державної реєстрації 0121U109211).

Внесок автора у результати наведених науково-дослідних робіт полягає в розробці інформаційної технології збору та оброблення структурованих та неструктурованих даних для моделювання та прогнозування досліджуваних соціально-економічних процесів.

Мета і завдання дослідження

Мета дисертаційної роботи — розроблення інформаційної технології, основу якої складають методи, моделі, алгоритми, призначені для збору та обробки даних для прогнозування витрат на пенсії та соціальні допомоги, що підвищить ефективність рішень в управлінні соціальною сферою.

Для досягнення мети визначені такі завдання:

- дослідити сучасні підходи до збору даних у соціальній сфері та підготовки їх до подальшого використання з метою прогнозування витрат на пенсії та соціальні допомоги;
- сформулювати загальні вимоги до інформаційної системи збору та обробки даних для аналізу та прогнозування витрат на пенсії і соціальні допомоги;
- визначити методику побудови підсистеми збору та оброблення даних та її функціонування у складі Єдиної системи соціальної сфери України;
- розробити методику опрацювання текстової інформації щодо потреб населення у соціальному забезпеченні, представленої у зверненнях громадян та соціальних мережах;
- розробити методику моделювання контингенту одержувачів пенсій та допомог в умовах невизначеності, в тому числі, спричиненої військовим конфліктом;
- створити методику підготовки даних для побудови моделей та прогнозування витрат на пенсії та соціальні допомоги;
- розробити програмне забезпечення для збору, оброблення, аналізу, прогнозування та візуалізації результатів прогнозування витрат на пенсії та соціальні допомоги на основі розроблених методів, виконати її апробацію та аналіз результатів.

Об'єкт дослідження – процеси збору та оброблення даних у соціальній сфері, пов'язані із фінансуванням витрат на пенсії та соціальні допомоги.

Предмет дослідження – методи, моделі та інформаційні технології збору та оброблення даних для визначення поточних та майбутніх витрат на пенсії та соціальні допомоги.

Методи дослідження. У роботі застосовано міждисциплінарний підхід, що поєднує методи системного аналізу, інтелектуального аналізу даних, математичного моделювання, машинного навчання. Для дослідження систем пенсійного забезпечення, соціального захисту і соціального забезпечення в умовах невизначеності та для узагальнення теоретичних і методологічних засад збору даних використано методи діалектичного пізнання, системного аналізу, індукції та дедукції, аналізу і синтезу, аналогій і зіставлення, формалізації й моделювання, дослідження часових рядів, кластерного та інтелектуального аналізу даних, сценарний підхід та елементи теорії прийняття рішень. Для збору та попередньої обробки структурованих і неструктурованих даних застосовано ETL-процеси, веб-скрейпінг, REST/OData API-запити, методи системного, інтелектуального й статистичного аналізу даних, експертне оцінювання, а також методи нормування, трансформації та обробки текстових даних. Для побудови моделей і формування прогнозів використано регресійний аналіз, методи аналізу часових рядів, ймовірісно-статистичне та економіко-математичне моделювання, дерева рішень і багатовимірні розподіли. Для розроблення методики застосування розроблених інформаційних технологій – методи й засоби проєктування та розроблення програмного забезпечення для інформаційно-аналітичних систем, включно з прикладним і web-програмуванням.

Наукова новизна одержаних результатів

вперше:

озроблено методологію збору та оброблення даних, що описують процеси у соціальній сфері, яка відрізняється робастністю результатів, що дозволяє підвищити ефективність бюджетного процесу як Пенсійного фонду України, так і державного та місцевих бюджетів,

озроблено новий метод побудови моделей та ансамблів моделей для визначення контингенту одержувачів пенсій та соціальних допомог, який відрізняється модифікованою комплексною структурою та високою адекватністю, що дозволяє підвищити якість управлінських рішень, що приймаються у соціальній сфері,

творено інформаційну технологію прогнозування витрат на пенсії та соціальні допомоги, в основу якої покладено поєднання принципів системного аналізу, методів обробки структурованих та неструктурованих даних, оцінювання якості даних, яка забезпечує високу якість прогнозів;

удосконалено:

- підхід до збору даних з державних реєстрів, відкритих API, даних НУО, стрім-потоків та веб-скрапінг, який забезпечує повне покриття різнопланової інформації (платежі, демографія, гуманітарні гранти) та підвищує репрезентативність вибірок для моделювання,

- розроблено метод моделювання соціально-економічних процесів, що мають місце у соціальній сфері в умовах невизначеності даних, який відрізняється від відомих урахуванням різних типів невизначеностей, що підвищує адекватність моделей і якість оцінок прогнозів за лінійними та нелінійними моделями;

отримали подальший розвиток:

истемна методологія проектування і реалізації систем підтримки прийняття рішень, яка відрізняється комплексним підходом до розв’язання задач підготовки даних, оцінювання структури і параметрів математичних моделей досліджуваних процесів з використанням множин критеріїв якості, і забезпечує високу якість проміжних та остаточних результатів моделювання, прогнозування і прийняття управлінських рішень;

метод прогнозування, заснований на використанні адаптивного підходу до моделювання у поєднанні із статистичним та ймовірнісним моделюванням, що дає можливість урахувати структурно-параметричні невизначеності і забезпечує адекватний опис причинно-наслідкових зв’язків і можливих варіантів розвитку процесів різної природи під впливом груп внутрішніх та зовнішніх чинників.

Практичне значення отриманих результатів полягає в тому, що на основі запропонованих в роботі інформаційних технологій, моделей, методів, алгоритмів збору та оброблення даних розроблено методику прогнозування витрат на пенсії і соціальні допомоги та реалізоване відповідне програмне забезпечення. Запропонована методика та програмне забезпечення призначені для використання у Пенсійному фонді України у складі ІКІС ПФУ та у Єдиній інформаційній системі соціальної сфери України.

Використання у соціальній сфері запропонованих в роботі інформаційних технологій, забезпечить підвищення якості управлінських рішень щодо реалізації соціальних гарантій в умовах невизначеності, в тому числі, спричиненої військовим станом. Запропоновані в роботі алгоритми збору, очищення та уніфікації даних є універсальними, вони призначені для розв’язання широкого кола актуальних задач інтеграції даних різних типів, включаючи роботу з персональними даними громадян, державними реєстрами, відкритими даними, API-платформами, даними з соціальних мереж та зверненнями громадян, отримання нових відомостей про об’єкт дослідження та виявлення закономірностей розвитку досліджуваних процесів. Це розширить аналітичні можливості системи, підвищити якість аналітичної роботи у соціальній сфері, в тому числі, у процесі формування та моніторингу виконання державного та місцевих бюджетів, бюджету Пенсійного фонду України. Запропонована у роботі методологія опрацювання даних та подальшого їх використання у єдиному інтегрованому інформаційному просторі для прогнозування витрат на виплату пенсій та соціальних допомог закладає основи для розвитку нового підходу до створення інформаційних систем, використовуваних у державному управлінні та місцевому самоврядуванні. Практична значущість результатів дослідження підтверджена довідками Головного управління Пенсійного фонду України у Львівській області, ТОВ «МЕДИРЕНТ».

Особистий внесок здобувача. Усі теоретичні та практичні результати, що виносяться на захист, отримані автором самостійно. Пошук та аналіз літературних джерел за тематикою дисертаційного дослідження виконано автором особисто. Наукові положення, висновки, рекомендації, що викладені в дисертації, розроблені особисто здобувачем. У працях, виконаних у співавторстві, автору дисертації належать: методика формування критеріальної бази та підготовки даних для аналізу, побудови моделей та прогнозування видатків на соціальний захист та соціальне забезпечення [1, 2, 6]; метод оброблення великих обсягів даних для побудови ймовірісно-статистичних моделей та ансамблів моделей для прогнозування динаміки та структури контингенту отримувачів пенсій та соціальних допомог [2, 3]; архітектура системи підтримки прийняття рішень для аналізу та прогнозування витрат на пенсії та соціальні допомоги [4]; функціональна схема та програмне забезпечення системи підтримки прийняття рішень із використанням запропонованої інформаційної технології [4, 9]. системна методологія використання текстової інформації для прогнозування видатків на соціальний захист та соціальне забезпечення [5, 7, 8].

Апробація результатів дисертації. Результати дисертаційного дослідження були представлені на 3-х міжнародних та всеукраїнських науково-практичних конференціях: XXII Міжнародній науково-практичній конференції «Інформаційно-комунікаційні технології та сталий розвиток» (Київ, 14-15 листопада 2023 р.), XXIII Міжнародній науково-практичній конференції «Математичне моделювання та інформаційно-комунікаційні технології для зміцнення та відновлення» (Київ, 12-13 листопада 2024 р.), VIII Міжнародній науково-практичній конференції «Прикладні інформаційні системи та технології в цифровому суспільстві» (Київ, 01 жовтня 2024 р.).

Публікації. За результатами дослідження опубліковано 8 наукових праць, у тому числі:

- 3 статті у наукових періодичних виданнях, включених до Переліку наукових фахових видань України (в т.ч. 1 - включених до категорії «А»);
- 2 статті у наукових періодичних виданнях інших держав з напрямку, з якого підготовлено дисертацію, що включені до міжнародних наукометричних баз SCOPUS та/або Web of Science Core Collection;
- 3 тези та доповіді на науково-практичних конференціях.

Одне авторське свідоцтво на комп'ютерну програму.

Якість та кількість публікацій відповідають «Вимогам до опублікування результатів дисертацій на здобуття наукових ступенів доктора і кандидата наук».

Структура та обсяг дисертації. Дисертація складається з вступу, чотирьох розділів, висновків та додатків. Вона містить 167 сторінок машинописного тексту, включаючи 30 рисунків, 34 таблиці, список використаних джерел з 152 найменувань.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтована актуальність досліджень, сформульовані мета і завдання досліджень, приведені основні наукові положення та результати, винесені на захист, а також відомості про практичне значення результатів роботи, їх апробацію і публікацію матеріалів дослідження.

У першому розділі представлено огляд сучасних інформаційно-аналітичних систем, що використовуються у державному та публічному управлінні в Україні та за кордоном, і, зокрема, у системі соціального захисту та соціального забезпечення. В результаті проведеного аналізу виявлено, що більшість з них зосереджені переважно на наданні доступу до даних державних реєстрів та баз даних. Аналітична складова або відсутня, або обмежена функціями моніторингу, формування звітності та планування на наступний рік. З-поміж інших вирізняється аналітична система, реалізована у складі АРМ керівника Інтегрованої комплексної інформаційної системи Пенсійного фонду України (ІКІС ПФУ). Дана система відрізняється наявністю інструментів Oracle BI, які використовуються переважно для формування звітності і візуалізації результатів. Схема моніторингу та аналітики у складі АРМ керівника представлена на рис.1.

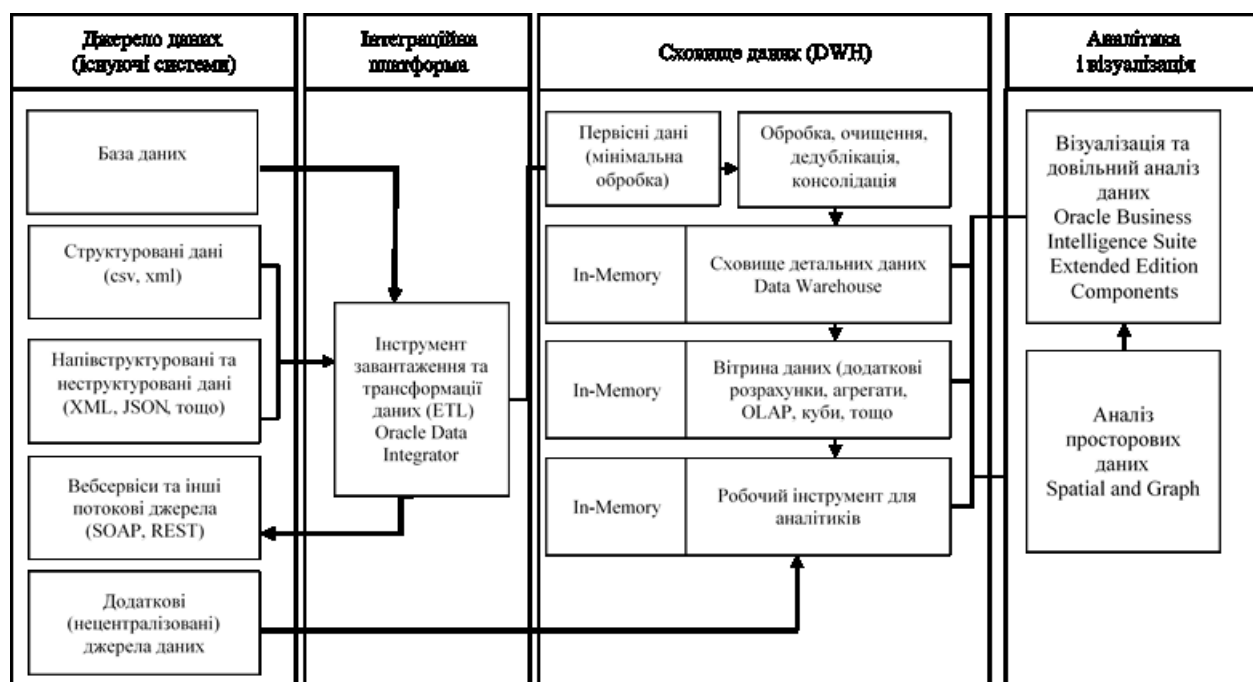


Рисунок 1 – Схема оброблення даних у модулі моніторингу та аналітики (АРМ Керівника) Пенсійного фонду України на платформі Oracle BI

Починаючи з 2020 року на базі ІКІС ПФУ створюється Єдина інформаційна система соціальної сфери України. Формується єдиний інформаційний простір - замість розрізнених реєстрів та баз даних впроваджується цілісна система з єдиною програмно-технічною архітектурою, централізованою базою даних та базою знань, здатна ефективно адаптуватись до нових задач та викликів, що постають перед соціальною сферою країни. Її складовими є загальна та прикладні підсистеми, працює функціональне ядро системи на базі якого здійснюється автоматизація прикладних задач. На даний час ще не завершено формування єдиної бази даних, однак найбільші бази даних – база даних Пенсійного фонду України та база даних одержувачів соціальних допомог та субсидій вже об'єднані. Триває розроблення загального інтерфейсу, налагодження системи обміну інформацією між її компонентами, створення єдиної системи кіберзахисту, тощо.

На основі системного дослідження предметної області, аналізу робіт вітчизняних та закордонних науковців та фахівців-практиків, встановлено, що для прогнозування витрат на виплату пенсій та соціальних допомог, крім даних, наявних у соціальній сфері, потрібно залучити дані з зовнішніх джерел. Для вирішення цієї задачі розроблена відповідна методика (рис. 2).

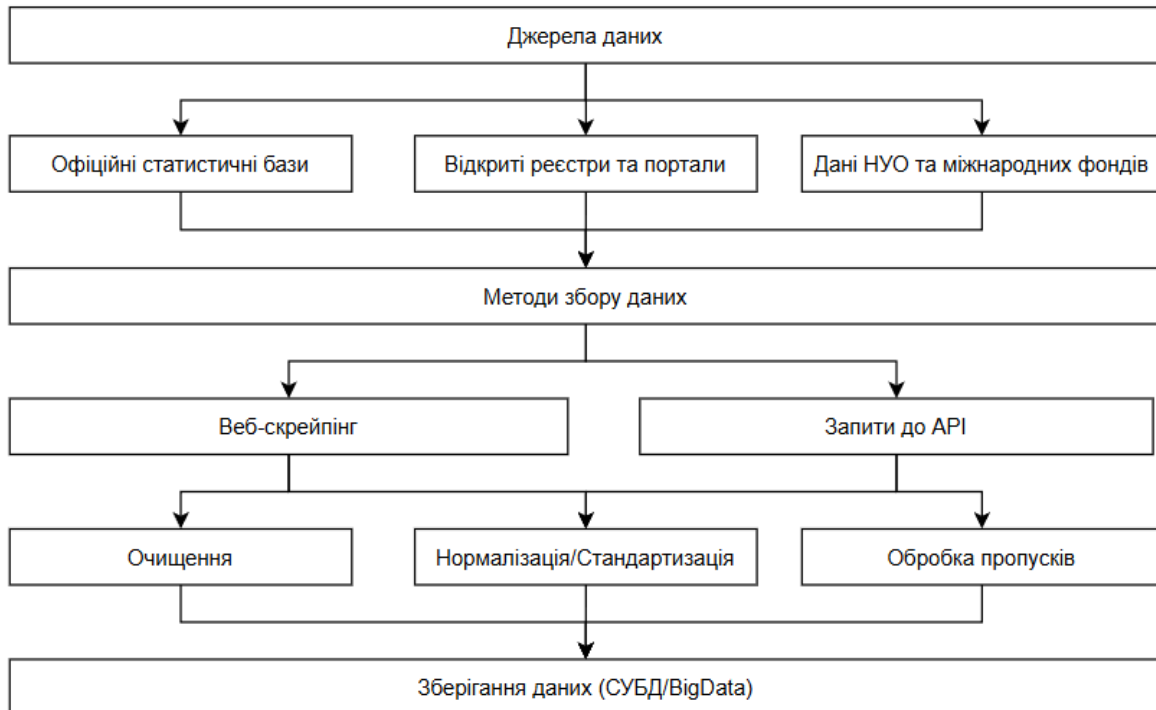


Рисунок 2 – Методика збору та попередньої обробки даних із зовнішніх джерел

Розроблений інтегрований підхід дозволяє поєднати у єдиній системі різні за типом, структурою та форматом представлення, дані, що наявні у державних реєстрах, аналітичних звітах, публікаціях Державної служби статистики, Національного банку України, відкритих API, Інтернет-сторінках міжнародних організацій, стрім-потоків, веб-скрапінгу, зверненнях громадян, публікації у соціальних мережах, інші текстові повідомлення та документи, що відображають потребу населення у соціальних послугах та допомогах, дають можливість урахувати настрої різних суспільних груп, а також поточних даних, що охоплюють інформацію від сенсорних систем, онлайн-моніторингу та дані, отримані від неурядових організацій, які характеризують оперативні зміни у середовищі дослідження. Запропонована методологія відрізняється робастністю результатів, що забезпечує репрезентативність вибірок та достовірність подальших прогнозів. Для цього удосконалено інструментарій ETL-процесів, що дозволяє здійснювати видобування, трансформацію та завантаження даних із різномірних джерел, розширено можливості веб-скрейпінгу та REST/OData API-запитів для інтеграції з відкритими платформами, а також реалізовано механізми потокової обробки даних у режимі реального часу.

Застосування розробленої методики на етапі системологічного аналізу проблеми забезпечує якнайповніше покриття предметної області різноплановою інформації (наприклад, видатки державного бюджету на соціальні виплати,

демографічна статистика, гуманітарні гранти, обсяги фінансування окремих соціальних програм з місцевих бюджетів, тощо), що підвищує репрезентативність вибірок для моделювання, покращуючи якість отримуваних прогнозів.

У другому розділі представлено результати аналізу існуючих методів та моделей збору та оброблення даних, на основі яких розроблено методику формування консолідованих даних, особливістю якої є інтегроване використання платформи Oracle та програмного забезпечення компанії SAS Institute (рис. 3).



Рисунок 3 - Методика формування консолідованих даних

Запропонована інформаційна технологія збору та оброблення даних, призначена для функціонування у складі Єдиної інформаційної системи соціальної сфери, яка функціонує на програмно-технічній платформі, що має трирівневу архітектуру, побудовану на основі системи керування базами даних Oracle Database Enterprise Edition 12, технології ASP.NET та веб-клієнта, для збору та обробки великих обсягів структурованої та неструктурованої інформації використовується Oracle Business Intelligence (Oracle BI).

Для порівняння продуктивності ключових інструментів збору даних — веб-скрейпінгу (BeautifulSoup, Scrapy, Selenium) та API-запитів (REST/OData до data.gov.ua і openbudget.gov.ua) в умовах реального навантаження було виконано тестування (умови проведення тесту: комп'ютер: 4-ядерний сервер, 8 GB RAM, інтернет-з'єднання 100 Mbps, час на кожний запуск: 1 година, одночасно запуснені по одному процесу кожного інструмента, вимірюємо: кількість успішно отриманих рядків (записів) за годину і число помилок HTTP 429 чи 5xx.)

Таблиця 1 – Результати експериментів з оброблення даних

Інструмент	Джерело	Записи/ годину	429/5xx помилки	Деталі реалізації
BeautifulSoup	HTML (treasury)	8 500	5×429	Синхронний підхід: <i>requests.get()</i> + <i>bs4</i> парсинг, без кешування запитів
Scrapy	HTML (treasury)	15 200	1×429	Асинхронна обробка черги запитів, вбудована пагінація, автокешування
Selenium	HTML (treasury)	2 100	0	WebDriver очікує рендерингу JS; імітація кліків; обмежена пропускна здатність
REST (Python requests)	data.gov.ua	30 000	0×429, 0×5xx	Прямі GET-запити з пагінацією; простий back-off при 429
OData (pandas)	openbudget.gov.ua	28 500	2×429	Завантаження JSON-контенту через <i>read_json()</i> ; автоматичне розбиття на ланки

Як показали результати експериментів (табл. 1), Scrapy показує найвищу пропускну здатність завдяки неблокуючому I/O та кешуванню HTTP-відповідей. REST/OData-запити до SKAN/OpenBudget API стабільні та майже не дають помилок сервера, але потребують обробки пагінації й ліміту запитів.

Для оброблення текстових даних, у роботі представлена методика, яка базується на методах лінгвістичного аналізу, кластеризації та інтелектуального аналізу даних. Вона забезпечує можливість виділення ключових тем і проблемних аспектів у зверненнях громадян, що суттєво підвищує релевантність і оперативність управлінських рішень. Розроблені методи текстової аналітики та кластеризації звернень населення дозволяють оперативно виявляти пріоритетні проблеми та потреби різних соціальних груп, що сприяє ухваленню обґрунтованих управлінських рішень.

Для розроблення аналітичної складової обрано програмне середовище SAS Institute (мова програмування SAS Base), яке добре інтегрується із апаратно-програмною платформою Oracle, і за своїми характеристиками є оптимальним для розв'язання аналітичних задач, має програмний інтерфейс SWAT API який дозволяє використовувати всі переваги платформи SAS Viya та мов програмування Open Source, зокрема, R та Python. Дані для аналізу використовуються не лише з аналітичного сховища (застосована інтеграція Hadoop та ETL/ELT), а й в результаті виконання API-запитів, в тому числі, коли потрібно здійснювати обробку у режимі реального часу (використано SAS Event Stream Processing). Крім того, використання функціоналу інструменту SAS Data Quality дозволить покращити якість вхідних даних, зокрема здійснити їх профілювання, перевірку, очищення та стандартизацію, що в свою чергу підвищить якість побудованих моделей. Гнучкість та адаптивність інформаційних технологій досягається за рахунок використання конвергентного підходу, який передбачає, що розроблювана система має центр обробки даних

(Big Data), систему збору, накопичення, обробки та завантаження даних у сховище або хмару, програмні засоби для математичного моделювання, інтелектуального аналізу даних, засоби штучного інтелекту, систему обміну даними між компонентами Єдиної інформаційної системи соціальної сфери України, взаємодії між користувачами різних рівнів та забезпечення надійності віддаленого доступу, систему інформаційної та кібербезпеки, зокрема, захисту персональних даних, систему моніторингу та адміністрування роботи системи, тощо.

У **третьому розділі** дисертаційної роботи зосереджено увагу на розробці та обґрунтуванні методів і моделей прогнозування витрат на пенсії та соціальні допомоги. Для побудови моделей запропоновано удосконалену методику покрокової структурно-параметричної адаптації моделей (рис 4).

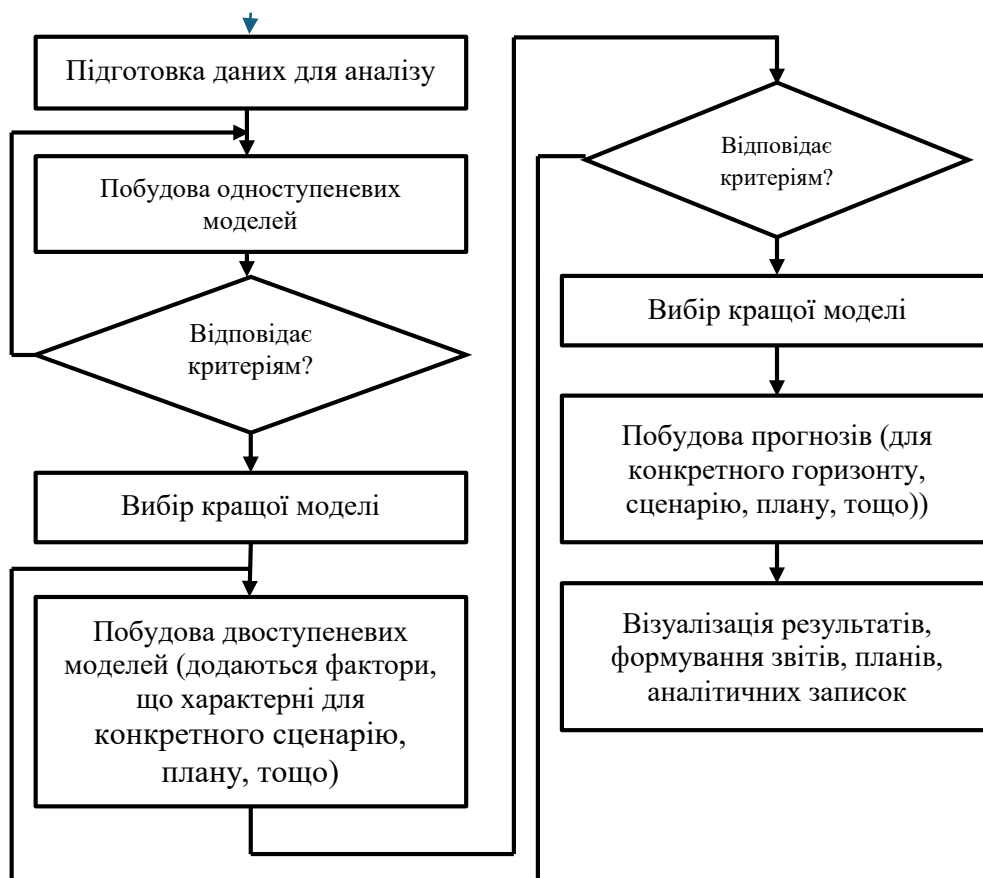


Рисунок 4 - Схема побудови і оцінювання прогнозів із використання методики структурно-параметричної адаптації моделей (фрагмент)

Враховано, що процедури збору, оброблення та зберігання даних відбуваються відповідно до встановлених регламентів, а дані мають бути захищені відповідно до вимог. Збір даних, використовуваних для побудови моделей, здійснюється у двох напрямках: індивідуальні дані застрахованих осіб і осіб, що потребують соціальної допомоги та соціальних послуг і узагальнена інформація стосовно потреби та реальних витрат на соціальний захист та соціальне забезпечення.

Спочатку будується множина одноступневих моделей, кожна з яких має період навчання та тестування, краща модель обирається на основі декількох

статистичних критеріїв оцінювання якості. Далі визначається краща модель, аналізуються залишки моделі та будуються двоступеневі моделі (або комбіновані моделі), серед яких обирається краща, яка буде використана для побудови прогнозу. Передбачено, що користувач має можливість додати потрібний елемент до технологічного ланцюга самостійно, використавши відповідні візуальні компоненти, як це реалізовано у середовищі SAS Enterprise Guide. Запропонована методика адаптивного прогнозування передбачає автоматичне оновлення параметрів моделей у процесі надходження нових даних. Поєднання статистичних та ймовірнісних методів дозволяє моделі «підлаштовуватися» під зміни у соціально-економічному середовищі.

Враховуючи особливості предметної області, в роботі запропоновано методику побудови ансамблів моделей, яка відрізняється модифікованою комплексною структурою та високим рівнем адекватності. Запропонований метод ансамблювання базується на поєднанні моделей декількох класів (табл. 2).

Таблиця 2 – Приклади ансамблювання моделей

Призначення	Клас моделей	Призначення
Інтелектуальний аналіз даних	регресійні моделі	виявлення основних трендів у динаміці показника
Побудова прогнозу	моделі часових рядів	врахування сезонності та нестационарності процесів
	ймовірнісно-статистичні	опис невизначеностей
Підвищення якості прогнозу	методи машинного навчання	підвищення точності класифікації та прогнозування

Такий підхід є актуальним у задачах, коли відбувається, наприклад зміна методики формування статистичних показників, що використовуються для побудови моделей, наприклад, використання рекурсивної регресії та методів ковзного вікна для оперативного оновлення прогнозів витрат Пенсійного фонду України на виплату пенсій залежно від змін у чинному законодавстві щодо умов нарахування пенсій певного виду за останній місяць.

Як показали чисельні експерименти, застосування розробленої методики, наприклад, ансамбль моделей складається з моделі ARIMA (для базового опису процесу) та Random Forest (для уточнення моделі), коли до моделі додаються додаткові регресори, дає можливість покращити прогнози динаміки витрат на виплату субсидій у окремих регіонах.

Для урахування невизначеностей різної природи (наприклад, відсутність точних даних щодо окремих категорій населення у певному регіоні), вплив внутрішніх (зміни у законодавстві, демографічні кризи, міграція) та зовнішніх чинників (військові дії, екологічні фактори, економічні кризи, ансамблі моделей) запропоновано використовувати у поєднанні з стохастичними сценаріями, які дозволяють оцінити діапазон можливих значень цільової змінної замість єдиного прогнозного показника. Такий підхід дозволив отримати прогнози високої якості у таких задачах, як прогнозування витрат на виплату соціальних допомог у регіонах за урахування внутрішньої міграції населення, внаслідок війни.

Четвертий розділ присвячений апробації результатів дослідження. В даній роботі представлені результати використання запропонованих технологій збору та оброблення даних та розробленого програмного забезпечення, які реалізують технологічний ланцюг, призначений для вирішення задачі прогнозування витрат на виплату пенсій та соціальних допомог.

Приклад 1. Прогнозування зміни контингенту одержувачів пенсій та соціальних допомог за рахунок бойових санітарних та безповоротних втрат. Для побудови прогнозу використано дані з відкритих джерел щодо кількості бойових санітарних та безповоротних втрат у період 2014-2019 рр. у зоні АТО. На їх основі було зроблено припущення стосовно співвідношення різних видів бойових втрат (рис. 5).

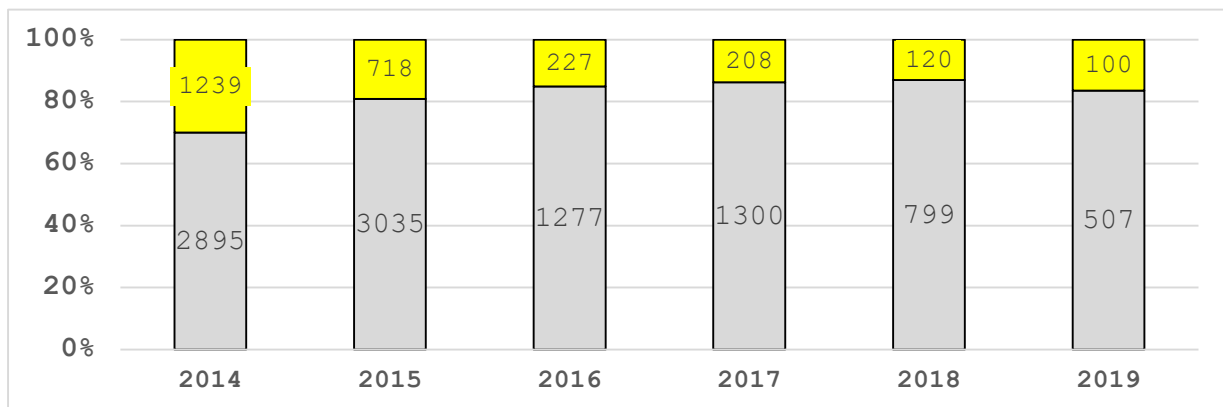


Рисунок 5 - Співвідношення бойових санітарних втрат (на діаграмі - світлий колір) і бойових безповоротних втрат (на діаграмі темний колір)

Як видно з рис. 5, санітарні втрати становлять переважну більшість втрат, їх кількість коливається залежно від інтенсивності та характеру бойових дій, захисту особового складу – найбільші безповоротні втрати були саме на початку АТО, коли ще не було напрацьовано бойового досвіду. Фактори, вплив яких на кількість санітарних втрат є найсуттєвішим, відібрані за даними експертів. Вказані фактори й були розглянуті під час моделювання санітарних втрат. На рис. 6 наведена топологія побудованої мережі Байєса.

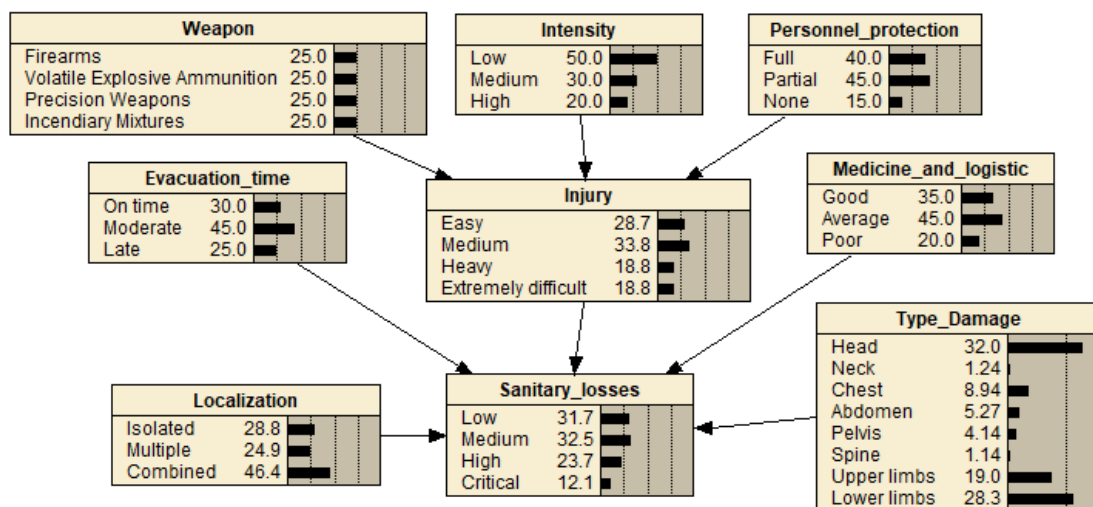


Рисунок 6 – Графічне представлення мережі Байєса для прогнозування ймовірності санітарних втрат

Для побудови мережі Байєса використані такі фактори, як «Інтенсивність бою», «Тип зброї», «Ступінь тяжкості поранення», «Захист особового складу», «Час евакуації», «Мед. забезпечення та логістика», «Характер ушкоджень», «Локалізація ушкоджень», «Тип санітарних втрат». Використовуючи ймовірісно-статистичні моделі, такі як представлено на рис. 6, розглянуто декілька сценаріїв зміни кількості санітарних втрат (табл. 3).

Таблиця 3 – Приклад сценаріїв моделювання із використанням наведеної мережі Байєса

Інтенсивність бою	Тривалість евакуації	Якість медичного забезпечення та логістики	Санітарні втрати, %			
			Низькі (< 1,5)	Середні (1,5 – 3)	Високі (3 – 5)	Критичні (> 5%)
Мала	Своєчасна	Якісне	75	20	5	0
Середня	Помірна	Середнє	30	45	20	5
Висока	Запізно	Погане	5	20	40	35
Середня	Запізно	Погане	10	30	40	20
Висока	Своєчасна	Якісне	20	45	25	10
Мала	Помірна	Середнє	50	35	12	3

За позитивного сценарію (табл. 3), коли мала інтенсивність бою, евакуація здійснюється вчасно, хороша якість медичного забезпечення та логістика, санітарні втрати будуть найменшими. За найгіршого сценарію, найвища ймовірність зростання кількості санітарних втрат до високих і, навіть, критичних значень. Змоделювати ситуацію можна, додавши вершину, яка відображає ймовірність настання інвалідності. На рис. 7 наведена топологія побудованої мережі Байєса, яка моделює ситуацію, за якої можлива інвалідність військового.

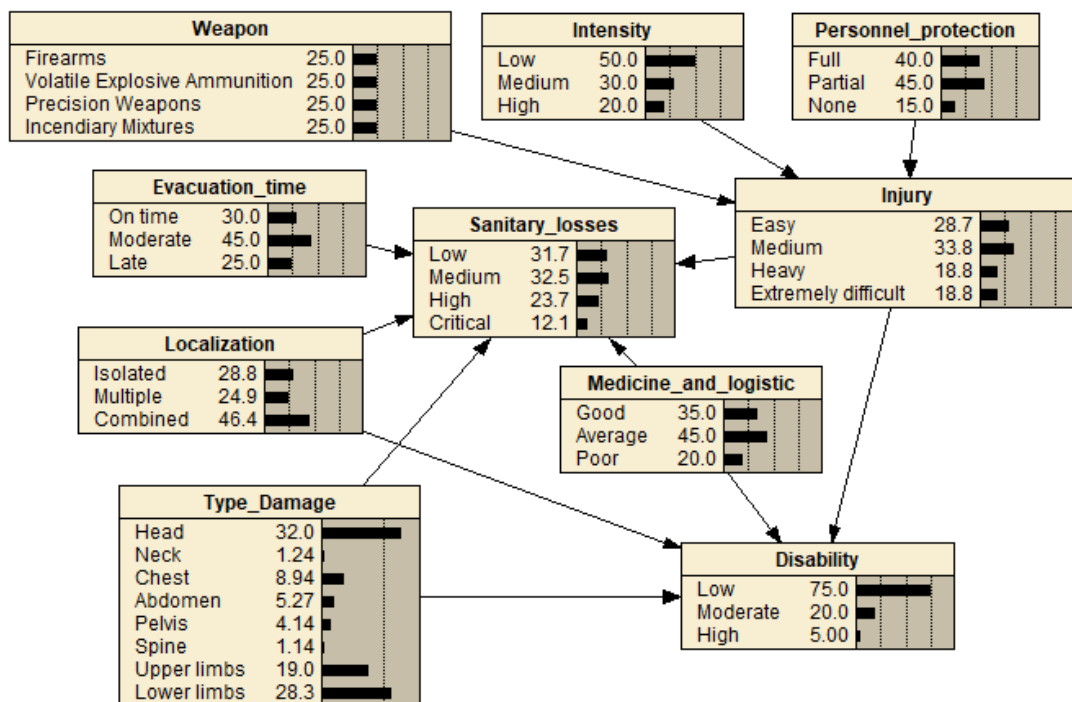


Рисунок 7 - Мережа Байєса для прогнозування ймовірності настання інвалідності

Таблиця 4 – Розподіл значень «Ймовірність отримати інвалідність» (Disability), для вершини мережі Байєса

Ймовірність отримати інвалідність	Позначення ймовірності отримання інвалідності в моделі	Значення ймовірності, %
Низька ймовірність	Low	75
Помірна ймовірність	Moderate	20
Висока ймовірність	High	5

Опис станів вершини «Ймовірність отримати інвалідність»

1) Низька ймовірність – як правило це випадки коли поранення не є критичними або особа має достатньо високий рівень медичного забезпечення, ефективний захист та швидку евакуацію.

2) Помірна ймовірність – зазвичай це серйозні поранення, але не критичні. Може бути затримка в евакуації або недостатня медична допомога, але шанс на повне відновлення залишається.

3) Висока ймовірність – коли поранення надзвичайно тяжкі, і ймовірність інвалідності дуже висока через серйозні ушкодження внутрішніх органів, ампутацію, тяжкі черепно-мозкові травми тощо. Окрім цього дуже суттєво впливають проблеми з евакуацією або з медичним забезпеченням.

Таблиця 5. Результати моделювання із використанням пропонованої ймовірнісно-статистичної моделі - мережі Байєса

Локалізація ушкоджень	Медичне забезпечення та логістика	Тип ушкодження	Характер ушкоджень	Ймовірність отримати інвалідність
Ізольовані	Якісне	Легка	Верхні кінцівки	Низька
Множинні	Середнє	Середня	Живіт	Помірна
Поєднанні	Погане	Вкрай важка	Голова	Висока
Ізольовані	Якісне	Легка	Верхні кінцівки	Низька
Множинні	Середнє	Середня	Живіт	Помірна
Поєднанні	Погане	Важка	Голова	Висока
Множинні	Середнє	Важка	Нижні кінцівки	Помірна
Поєднанні	Середнє	Вкрай важка	Хребет	Висока

В результаті моделювання отримано такі сценарії: Перший сценарій – легкі ушкодження (наприклад, пошкодження верхніх або нижніх кінцівок), швидке медичне забезпечення, ізольований і легкий тип ушкодження. У такому випадку ймовірність інвалідності буде низькою. Другий сценарій передбачає, що особа може отримати середні ушкодження, ізольовані або множинні і висока ймовірність затримки медичної допомоги. За таких умов, ймовірність інвалідності зростає до помірного рівня. За третього сценарію особа отримує вкрай тяжкі ушкодження (пошкодження голови або хребта), ситуація ускладнюється відсутністю швидкого надання медичної допомоги. В такому випадку ризик інвалідності дуже високий. Отже, як показали проведені чисельні експерименти, запропонований підхід може бути використаний для

прогнозування санітарних втрат, і на основі отриманого прогнозу – кількості поранених, які з високою ймовірністю будуть інвалідами.

Приклад 2. Застосування інтелектуального аналізу текстової інформації у задачі виявлення потреб населення у соціальній допомозі та соціальних послугах. Вхідною інформацією для дослідження є електронні звернення громадян, які надійшли до вебпорталу електронних послуг Пенсійного фонду України та державної установи «Урядовий контактний центр». Були досліджені й матеріали Інтернет-видань, різних за тематикою та аудиторією, державних та недержавних. Джерела за тематикою та популярністю були згруповані у 8 кластерів. На основі аналізу текстів, що стосуються питань соціального захисту та соціального забезпечення, розміщених на вказаних Інтернет-ресурсах та у електронних зверненнях, було отримано шість кластерів. На основі попереднього аналізу текстів звернень сформовано корпус текстів, фрагмент якого наведений в табл. 6.

Таблиця 6 - Частотна матриця термів для корпусу текстів, побудована на основі корпусу текстів, сформованого із електронних звернень громадян

Позначення	Терм	Кількість згадувань у документі:									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
t1	суд	1	0	0	0	0	0	1	2	0	0
t2	надбавки	1	0	1	1	0	0	1	0	2	0
t3	військово-службовці	0	1	0	0	2	1	0	0	0	0
t4	грошове забезпечення	0	1	0	0	1	0	0	0	2	0
t5	пенсія	0	1	0	1	2	2	1	0	1	1
t6	правоохоронців	0	1	0	0	1	0	0	0	0	0
t7	колишніх	0	1	0	0	1	0	0	0	0	0
t8	аварія	0	0	1	1	0	0	0	0	0	1
t9	ЧАЕС	0	0	1	2	0	0	0	0	0	1
t10	Україна	0	0	1	0	0	1	0	0	0	0
t11	отримувала	0	0	0	1	1	0	0	0	0	0
t12	вислуга	0	0	0	0	1	1	0	0	0	0

Для вирішення проблеми зменшення розмірності та розрідженості частотної матриці S корпусу текстів було використано метод сингулярного розкладу (SVD). На основі отриманої матриці S можна обчислити яку кількість у відсотках вносить в пояснення даних відповідний вимір, що описується відповідним сингулярним значенням. За результатами аналізу отриманих сингулярних значень визначено, що два базових виміри, можуть пояснити 66,16% варіабельності даних. В такому випадку всі документи можна розташувати у двовимірному просторі і визначити кластери, які вони утворюють за ступенем подібності та приналежності до певної теми. Як показали розрахунки, перший вимір пояснює 45,61% варіабельності даних, другий вимір пояснює 20,55% варіабельності даних. В результаті сформовано три тематичних

кластери, до яких увійшли документи за подібністю використання термів. Для автоматизації оброблення текстових даних використано систему SAS Text Miner.

Статистичні характеристики побудованої моделі класифікації на основі лінгвістичних правил було обчислено окремо для тренінгового та тестового наборів даних: співвідношення - 70% для тренінгу та 30% для тестування, тобто 114 та 48 текстів відповідно. Результати узагальнені в табл. 7.

Таблиця 7 - Статистичні характеристики моделі класифікації досліджуваних текстів

Статистика	Набір даних	
	тренінговий	тестовий
TP (True Positive)	30	11
TN (True Negative)	67	26
FP (false positive)	10	6
FN (false Negative)	7	5
MISC, % (частка неправильно класифікованих значень)	15	23
GINI	0,82	0,71
ROC	0,79	0,67

Побудовані лінгвістичні правила було використано для кластеризації близько 10 тис. текстів новин за тематикою соціального захисту та соціального забезпечення, що були опубліковані в мережі Інтернет з вересня 2023 року по вересень 2024 року. Після кластеризації текстів, для кожного кластеру було розраховано кількість текстів, які належать дописувачам з певного регіону. Отримані значення було нормовано по шкалі від 0 до 100 за формулою:

$$popularity_i = \frac{n_i}{\max(n_i | \forall i)}$$

де $popularity_i$ – популярність текстів відповідного кластеру, для i -го регіону, n_i – кількість текстів по регіону, $\max(n_i | \forall i)$ – максимальна кількість текстів за всіма регіонами

Результати кластерного аналізу текстової інформації візуалізовано використовуючи інструменти SAS Enterprise Guide 7.1 (рис 8).

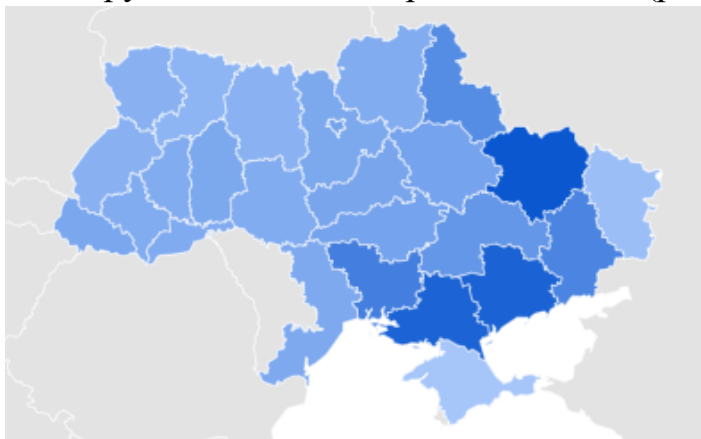


Рисунок 8 - Кластер 3, популярності текстів за темою «Проблеми пов'язані з ВПО» по регіонах України

Отже, запропонована методика дозволяє виявляти ситуації, які потенційно можуть вимагати заходів соціальної підтримки, що забезпечить адресність допомоги.

ВИСНОВКИ ТА ОСНОВНІ РЕЗУЛЬТАТИ РОБОТИ

В результаті виконання дисертаційної роботи отримано такі наукові результати.

1. Розроблено нові інформаційні технології збору та оброблення даних для прогнозування витрат на виплату пенсій та соціальних допомог. Їх основу становлять математичні моделі та їх ансамблі, методи інтелектуального аналізу даних, штучний інтелект, сучасні технології збору та зберігання даних, поєднання яких дозволяє накопичувати та обробляти різномірну інформацію, отриману з різних джерел та використовувати її для прогнозування витрат на пенсії та соціальні допомоги. Розроблені інформаційні технології є складовою інформаційно-аналітичної платформи, яка вирізняється застосуванням конвергентного підходу, інтеграцією аналітичних платформ SAS та Oracle BI на бази апаратно-програмного забезпечення компанії Oracle, гнучкістю та адаптивністю, можуть бути використані окремо або в складі інформаційно-аналітичних систем органів державного управління та публічного управління.

2. Розроблено методологію збору та оброблення даних, що описують процеси у соціальній сфері, яка відрізняється робастністю результатів, використанням нового циклу обробки даних - від асинхронного ETL-збирання та LOESS-імпутації до автоматичного підбору моделей та надання візуальної інформації. Це забезпечує безперервну підготовку, аналіз і прогноз соціальних видатків навіть за неповних або спотворених даних, підвищуючи якість прогнозів, які створюють основу для прийняття рішень при плануванні витрат бюджетів різних рівнів на соціальний захист та соціальне забезпечення.

3. Удосконалено підхід до збору даних з державних реєстрів, відкриті API, дані НУО, стрім-потoki та веб-скрапінг, який забезпечує повне покриття різнопланової інформації (платежі, демографія, гуманітарні гранти) та підвищує репрезентативність вибірок для моделювання.

4. Розроблено новий метод побудови моделей та ансамблів моделей для визначення контингенту одержувачів пенсій та соціальних допомог, який відрізняються модифікованою комплексною структурою, адаптивністю та високою адекватністю, за рахунок використання регресійних та сезонно-регресійних моделей (основний орієнтир для моделювання), деревні ансамблі, такі як градієнтний бустинг та випадковий ліс, що забезпечують високу точність короткострокових прогнозів, особливо із нелінійними збуреннями.

5. Удосконалено алгоритм динамічного вибору моделі залежно від вхідних даних. Розроблено метод моделювання соціально-економічних процесів, що мають місце у соціальній сфері в умовах невизначеності, який відрізняється від відомих опрацюванням різних типів невизначеностей, що підвищує адекватність моделей і якість оцінок прогнозів за лінійними та нелінійними моделями. Пропонований підхід дозволяє підвищити якість прогнозів видатків у соціальній сфері.

6. Розроблено інформаційну технологію прогнозування витрат на пенсії та соціальні допомоги, в основу якої покладено поєднання принципів системного аналізу, методів обробки структурованих та неструктурованих даних, оцінювання якості даних, яка забезпечує високу якість прогнозів. Для урахування структурних невизначеностей розроблено процедуру оновлення параметрів моделей у режимі «наближення до реального часу», що враховує зовнішні шоки (бойові дії, демографічні зрушення) використовуючи моделі ARIMAX/SARIMAX з екзогенними факторами, для візуалізації результатів інтегровано фронтенд-дашборди, що генерують регіональні та національні сценарії соціальних витрат, візуалізують довірчі інтервали й ранжують ризики перевитрат.

Використання розробленої інформаційно-аналітичної підсистеми забезпечить як скорочення часу на оброблення значних обсягів даних (структурованих та неструктурованих) про застрахованих осіб, одержувачів соціальних допомог, так і підвищення якості прогнозів потреби у фінансуванні пенсій та допомог, дозволить отримувати результати аналізу та прогнозування у режимі реального часу, що покращить якість управлінських рішень в соціальній сфері країни. В подальшому система може бути розширена за рахунок додавання нових математичних моделей та їх ансамблів, засобів обробки геопросторової інформації та інших методів штучного інтелекту.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті у періодичних виданнях, включених до категорії «А» Переліку наукових фахових видань України

1. Зарудний О. Б. і Коваль Р. Г. Методика прогнозування часових рядів для визначення потреби у видатках на соціальний захист та соціальне забезпечення», *Міжнародний науково-технічний журнал «Проблеми керування та інформатики»*. 2024. Vol. 69, No. 5. С. 64–77. <https://doi.org/10.34229/1028-0979-2024-5-5>

Статті у наукових фахових виданнях України

2. Zarudnyi, O., Koval, R. Application of probabilistic and stochastic models and data mining for forecasting the contingent of old age pension recipients in the context of systemic uncertainty. *Technology Audit and Production Reserves*, 2024. Vol. 5. No. 2(79), P. 56–62. <https://doi.org/10.15587/2706-5448.2024.313960>

3. Трофимчук О., Коваль Р., Зарудний О. Методика прогнозування кількості інвалідів з числа санітарних втрат. *Екологічна безпека та природокористування*, 2025. Vol. 54. No. 2. С. 121–135. <https://doi.org/10.32347/2411-4049.2025.2.121-135>

Статті у закордонних виданнях, індексованих у базі даних Scopus

4. Zarudnyi, O., Koval, R. Methodology for Constructing an Analytical Subsystem of the Unified Information System of the Social Sphere of Ukraine. *Cybernetics and Systems Analysis*. 2025. Vol. 61. No. 4. P. 487–494. <https://doi.org/10.1007/s10559-025-00785-9> (Scopus)

5. Korbicz J., Sholokhov O., Koval R., Zarudnyi O. Application of SAS Text Miner for the analysis of citizens' appeals in the system of social protection and social

security. *8th International Scientific and Practical Conference Applied Information Systems and Technologies in the Digital Society*. (AISTDS 2024). (Kyiv , 1 Oct. 2024). 2024. Vol. 3942, P. 113 – 124. ISSN 1613-0073 URL: <https://www.scopus.com/pages/publications/105001096518> (Scopus)

Наукові праці, які засвідчують апробацію матеріалів дисертації

6. Зарудний О.Б., Коваль Р. Г. Проблеми прогнозування факторів, що впливають на обсяг видатків Пенсійного фонду України Інформаційно-комунікаційні технології для перемоги та відновлення // Колективна монографія за матеріалами XXII Міжнародної науково-практичної конференції «Інформаційно-комунікаційні технології та сталий розвиток» (Київ, 14-15 листопада 2023 р.) / За заг. ред. С.О. Довгого. – К.: ТОВ «Видавництво «Юстон», 2023. С. 76-79. URL: https://itgip.org/wp-content/uploads/2023/11/1_zbirka_08_11_23-1-1.pdf

7. Зарудний О.Б., Коваль Р. Г. Застосування методу сингулярного розкладання у задачі аналізу електронних звернень громадян до Пенсійного фонду України. Математичне моделювання та інформаційно-комунікаційні технології для зміцнення та відновлення // Колективна монографія за матеріалами XXIII Міжнародної науково-практичної конференції «Математичне моделювання та інформаційно-комунікаційні технології для зміцнення та відновлення» (Київ, 12-13 листопада 2024 р.) / За заг. ред. С.О. Довгого. К.: ТОВ «Видавництво «Юстон», 2024. С. 156-159. URL: https://itgip.org/wp-content/uploads/2024/11/2024-11-24_zbirka_all_07_11_2024_148x210.pdf

8. Зарудний О. Б., Коваль Р. Г. Шолохов О. В. Методика застосування кластеризації текстів на основі лінгвістичних правил для дослідження потреб населення у соціальному захисті та соціальному забезпеченні Прикладні інформаційні системи та технології в цифровому суспільстві: зб. Тез доповідей і наук. повідомлень учасників VIII Міжнародної науково-практичної конференції «Прикладні інформаційні системи та технології в цифровому суспільстві» (Київ, 01 жовтня 2024 р.) // за заг. ред. В. Плескач, О. Фендьо. К.: Київський нац. ун-т ім. Тараса Шевченка, 2024. С. 167-178. URL: https://aistis.knu.ua/wp-content/uploads/2024/10/Text_Book_conference_01.10.2024.pdf

Авторське свідцтво

9. Зарудний О. Б., Колбун В. А., Малецький О. М., Машкін В. Г., Рябцева Т. Б., Матвеєв С. В., Хандрос Ю. О., Коваль Р. Г., Бондарчук О. М., Педан Ю. А., Павлюков І. О. Комп'ютерна програма «Інтегрована комплексна система Пенсійного фонду України «ІКІС ПФУ», версія 2» №11056; заявл.31.10.2011; опубл. 11.11.2011.

АНОТАЦІЯ

Коваль Р. Г. Інформаційні технології збору та оброблення даних для прогнозування витрат на пенсії і соціальні допомоги. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – «Інформаційні технології» – Інститут телекомунікацій і глобального інформаційного простору НАН України, Київ, 2025 р.

Дисертаційна робота присвячена дослідженню методів, засобів, технологій

збору та оброблення даних щодо процесів, які відбуваються у соціальній сфері країни, використання прогнозуючих моделей високого ступеня адекватності, які дозволяють забезпечувати високу якість прогнозів витрат на пенсії та соціальні допомоги.

Створено нові інформаційні технології збору та оброблення даних, моделювання та прогнозування витрат на виплату пенсій та соціальних допомог шляхом інтеграції моделей, методів, алгоритмів і програмних засобів, які забезпечують високу адекватність моделей і задану якість прогнозів, призначені до використання у Єдиній інформаційній системі соціальної сфери України, а також є простими у використанні та зручними для користувачів. Запропонована методика побудови моделі процесів, що відбуваються у соціальній сфері відзначається робастністю, адаптивністю, є гнучкою, призначена для реалізації у комп'ютерних системах підтримки прийняття рішень. Запропоновано модифіковану методику моделювання часових рядів, на основі якої побудовано ансамблі моделей і вдосконалено методику використання структурованих та неструктурованих даних у задачах прогнозування витрат на пенсії та соціальні допомоги.

Ключові слова: інформаційні технології, математичні моделі, соціальна сфера, прогнозування, штучний інтелект, збір та оброблення даних.

ABSTRACT

Koval R. G. Information technologies for collecting and processing data for forecasting expenses for pensions and social benefits. – qualification scientific work in the form of a manuscript.

Dissertation for the degree of candidate of technical sciences in the specialty 05.13.06 – “Information technologies” – Institute of telecommunications and global information space of the NAS of Ukraine, Kyiv, 2025.

The dissertation is devoted to the study of methods, means, technologies for collecting and processing data on processes occurring in the social sphere of the country, the use of predictive models with a high degree of adequacy, which allow ensuring high quality forecasts of expenses for pensions and social benefits.

New information technologies for data collection and processing, modeling and forecasting of pension and social assistance costs have been created by integrating models, methods, algorithms and software tools that ensure high adequacy of models and a given quality of forecasts, are intended for use in the unified information system of the social sphere of Ukraine, and are also easy to use and user-friendly. The proposed methodology for building a model of processes occurring in the social sphere is characterized by robustness, adaptability, is flexible, and is intended for implementation in computer decision support systems. A modified time series modeling methodology has been proposed, on the basis of which ensembles of models have been built and the methodology for using structured and unstructured data in the tasks of forecasting pension and social assistance costs has been improved.

Keywords: information technologies, mathematical models, social sphere, forecasting, artificial intelligence, data collection and processing.