

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ**

Кафедра статистики та економічного аналізу

Савчук В. К., Воляк Л. Р.

**Методичні вказівки до вивчення дисципліни «Дейтамайнінг» для
здобувачів третього (освітньо-наукового) рівня вищої освіти ОНП «Облік і
оподаткування» зі спеціальності 071 «Облік і оподаткування»**

Київ 2024

Рекомендовано Вченою радою економічного факультету Національного університету біоресурсів і природокористування України

Укладачі: д. е. н., професор Савчук В. К., к. е. н., доцент Воляк Л. Р.

Рецензенти: д.е.н., проф. Негода Ю.В., к.е.н., доц. Кузик Н.П.

Методичні вказівки до вивчення дисципліни «Дейтамайнінг» для здобувачів третього (освітньо-наукового) рівня вищої освіти ОНП «Облік і оподаткування» зі спеціальності 071 «Облік і оподаткування»

Укладачі: д. е. н., професор САВЧУК Василь Кирилович
к. е. н., доцент ВОЛЯК Леся Романівна

Відповідальний за випуск: к. екон. наук, доцент МУЗИЧЕНКО А. О.

Київ:
Видавничий центр
НУБіП України,
2024.

ЗМІСТ

Мета і завдання навчальної дисципліни

Основні питання лекційного курсу

План практичних занять

Навчально-методичне забезпечення

Контрольні заходи перевірки знань

Рекомендовані джерела інформації

Електронні ресурси

МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Метою викладання навчальної дисципліни «Дейтамайнінг» є вивчення методів сучасної обробки даних – інтелектуального аналізу даних (Data Mining, Knowledge Discovery in Data), аналітичного дослідження великих масивів інформації з метою виявлення нових раніше невідомих, практично корисних знань і закономірностей, необхідних для прийняття рішень; огляд методів, програмних продуктів і різних інструментальних засобів, які використовуються Data Mining; розгляд практичних прикладів застосування Data Mining; підготовка аспірантів до самостійної роботи з вирішення задач засобами Data Mining і розробки інтелектуальних систем.

Data Mining – мультидисциплінарна область, яка виникла і розвивається на базі таких наук як прикладна статистика, розпізнавання образів, штучний інтелект, теорія баз даних та ін.

Завданнями вивчення дисципліни «Дейтамайнінг» є:

- здійснення пошуку, аналіз та критичне осмислення інформації, отриманої з різних джерел;
- генерування і продукування власних науково-теоретичних та практично-спрямованих ідей;
- робота з різними джерелами, аналіз, систематизація отриманої інформації;
- опанування базовими принципами побудови моделей даних;
- ознайомлення з концепцією Knowledge Discovery in Data (виявлення знань в даних) і Data Mining («видобування» знань);
- навчитися ефективно використовувати методи здобуття знань з великих масивів даних;
- отримати практичні навички з використання інструментальних засобів інтелектуального аналізу даних при вирішенні прикладних задач та навчитися інтерпретувати отримані результати;
- проводити емпіричні дослідження та використовувати економіко-математичні методи для встановлення тенденцій розвитку об'єктів дослідження;
- застосовувати інформаційні технології у науковій діяльності, сучасні методи наукових досліджень із використанням новітніх прикладних пакетів і програмних продуктів для наукового обґрунтування та підтвердження / спростування висунутих гіпотез;
- ідентифікувати наукові та практичні проблеми, здійснювати апробацію результатів наукових досліджень, висновків і практичних рекомендацій.

Згідно з вимогами освітньо-наукової програми аспіранти повинні:

знати :

- систему та показники розвитку економічних систем, методи збирання, обробки та аналізу даних;
- базові методологічні засади аналізу даних;

- основні поняття, задачі та стадії інтелектуального аналізу даних;
- підходи до збереження, представлення та обробки інформації в сучасних інформаційних системах;
- методи побудови моделей та аналізу взаємозв'язків у великих масивах даних;
- сучасні програмні засоби для проєктування і розробки систем інтелектуального аналізу даних;
- концепції сховищ даних, їх оперативної аналітичної обробки для практичного використання;
- прикладні пакети аналізу для обробки великих масивів даних.

вміти :

- генерувати нові ідеї на підставі креативного мислення;
- здійснювати пошук, оброблення та аналіз інформації з різних джерел;
- оцінювати та забезпечувати якість дослідницької та наукової діяльності;
- працювати з різними інформаційними ресурсами, отримувати, обробляти, аналізувати, систематизувати інформацію;
- обґрунтовувати вибір конкретного типу моделі та методу інтелектуального аналізу даних при вирішенні поставленої практичної задачі;
- проводити необхідну попередню обробку даних, визначати тип задачі аналізу, вирішувати її адекватно з обраним методом та оптимально визначеними параметрами, оцінювати результати, робити змістовні висновки та інтерпретацію;
- виявляти та усвідомлювати економічні зв'язки і процеси, які мають загальнодержавний характер для типологічно однорідних умов (економічних систем, видів діяльності);
- проводити емпіричні дослідження та використовувати економіко-математичні методи аналізу для встановлення тенденцій розвитку об'єктів дослідження;
- застосовувати інформаційні технології, сучасні методи наукових досліджень із використанням новітніх прикладних пакетів і програмних продуктів для обґрунтування та підтвердження / спростування гіпотез;
- застосовувати технології роботи зі сховищами даних, здійснювати їх аналітичну обробку та інтелектуальний аналіз для забезпечення надійної роботи інформаційних систем;
- проєктувати інформаційне забезпечення (логічну та фізичну структури баз даних) інформаційних систем;
- здійснювати апробацію і впровадження результатів наукових досліджень, висновків, методичних та практичних рекомендацій.

Засвоєння дисципліни забезпечить здобувачів здатністю розв'язувати комплексні проблеми в галузі професійної, у тому числі дослідницько-інноваційної діяльності, що передбачає творче переосмислення та створення нових, цілісних знань та/або ефективної професійної практики.

Основною формою засвоєння аспірантами знань для виконання наукового дослідження є самостійна робота, яка передбачає опрацювання монографій, наукових статей, матеріалів науково-практичних конференцій, авторефератів, дисертацій, підручників, навчальних посібників, інших науково-навчально-методичних джерел, законодавства в сфері розвитку аграрної сфери економіки України, обліку і оподаткування та діджиталізації. Самостійна робота регламентується РНП і становить для дисципліни: 75 год. (денна форма навчання), 130 год. (заочна).

Згідно з навчальним планом і ОНП бюджет навчального часу для вивчення дисципліни «Дейтамайнінг» (для всіх форм навчання) становить 150 годин (5 кредитів), в т. ч.: денна форма навчання 30 год. лекцій, 45 год. практичних і 75 год. – самостійна робота; заочна форма навчання – 8 год лекцій, 12 год. практичних, 130 год. – самостійна робота.

Форма підсумкового контролю - екзамен, що враховує результати оцінки виконаних практичних завдань та підсумкової атестації.

Питання лекційного курсу

Навчальний процес згідно з робочою програмою навчальної дисципліни «Дейтамайнінг» здійснюється у формах: лекції та практичні заняття, самостійна робота здобувача, контрольні заходи. Вивчення здійснюється шляхом оволодіння лекційним матеріалом, викладеного за планом та виконання практичних завдань.

Лекція 1. Концепція дейтамайнінгу

Визначення і область застосування. Data Mining. Поняття Business Intelligence. Цикл одержання, попередньої обробки, аналізу даних, інтерпретації результатів та їхнього використання. Етапи процесу Data Mining, пов'язані з побудовою, перевіркою, оцінкою, вибором і корекцією моделей. Методи первісної обробки даних. Методи дослідження структури даних: візуалізація та автоматичне групування даних.

Ключові поняття

Дейтамайнінг (інтелектуальний аналіз даних, Data Mining) – це технологія виявлення прихованих взаємозв'язків (корисних фактів) усередині великих баз даних.

Business Intelligence (інтелектуальний аналіз даних бізнесу, бізнес-аналітика) — комп'ютерні методи та інструменти для організацій, що забезпечують переклад транзакційної ділової інформації в форму, придатну для бізнес-аналізу, а також засоби для роботи з обробленою таким чином інформацією.

Штучний інтелект - здатність систем оброблення даних виконувати функції, що асоціюються з інтелектом людини, такі як логічне мислення, навчання та самовдосконалення (*Державний Стандарт України ДСТУ 2938-94*). Тобто, властивість автоматизованих систем брати на себе окремі функції інтелекту людини (наприклад, вибирати і приймати оптимальні рішення на основі раніше отриманого досвіду і раціонального аналізу зовнішніх впливів).

Системи візуалізації - це системи, які можуть подібно людині візуально взаємодіяти зі своїм середовищем, використовувати візуальні зображення та слухові сигнали для інструктування комп'ютерів або інших пристроїв.

Системи з навчання містять низку операцій, які надають можливість комп'ютеру або іншому зовнішньому пристрою набувати нових знань на додаток до того, що було вже введено раніше в пам'ять фірмою-виробником або програмістами.

«Знання» (Knowledge) у системах підтримки прийняття рішень, зокрема орієнтованих на знання, має виключно важливе значення. У загальному тлумаченні термін «знання» означає те, що будь-хто щось знає та розуміє. Нагромаджені знання зберігаються в базах знань.

База знань - це низка фактів, правил і процедур, які організуються в систему за допомогою специфічних програмних засобів, що забезпечують

пошук, зберігання, перетворення й занесення в пам'ять ЕОМ структурованих одиниць знань.

Knowledge Discovery in Data – виявлення знань в даних.

Knowledge-based systems – це системи знань у яких використовуються людські знання для розв'язання проблем, зокрема в бізнесі. Найпопулярнішим типом таких систем є експертні системи.

Експертна система — це комп'ютерна програма, у якій намагаються подати знання людини-експерта у формі евристик.

Евристика (від давньогрецького εὐρίσκω – «відшукую», «відкриваю») – сукупність логічних прийомів, методів і правил, що полегшують і спрощують рішення пізнавальних, конструктивних, практичних завдань.

Питання для контролю знань

1. Сутність поняття Data Mining
2. Область застосування і задачі Data Mining
3. Методи і моделі Data Mining
4. Поняття Business Intelligence
5. Цикл одержання, попередньої обробки, аналізу даних, інтерпретації результатів та їхнього використання.
6. Етапи процесу Data Mining, пов'язані з побудовою, перевіркою, оцінкою, вибором і корекцією моделей.

Лекція 2. Алгоритми Data Mining: класифікація і регресія

Постановка задачі класифікації та представлення результатів. Методи побудови правил класифікації. Методи побудови математичних функцій. Класифікація об'єктів у випадку невідомих розподілень даних. Методи оцінювання помилок класифікації. Методи вирішення задач регресії.

Ключові поняття

Класифікація - це процес розподілу об'єктів або даних на різні категорії або класи залежно від їх характеристик або властивостей. Це широко використовується в машинному навчанні та іншій обробці даних для розв'язання завдань, таких як розпізнавання образів, прогнозування і багато інших.

Регресія (англ. regression, нім. Regression) — форма зв'язку між випадковими величинами. Розрізняють прямолінійну, криволінійну, ортогональну, параболічну та інші регресії, а також лінію і площину регресії.

Питання для контролю знань

1. Методи первісної обробки даних.
2. Візуалізація та автоматичне групування даних.
3. Постановка задачі класифікації та представлення результатів
4. Методи побудови правил класифікації.
5. Методи побудови математичних функцій

6. Класифікація об'єктів у випадку невідомих розподілень даних.
7. Методи оцінювання помилок класифікації
8. Методи вирішення задач регресії.

Лекція 3. Інтелектуальний аналіз часових рядів

Поняття нечітких часових рядів. Методи моделювання часових рядів. Методи аналізу та прогнозування поведінки часових рядів.

Ключові поняття

Часовий ряд – це послідовність впорядкованих у часі числових показників, що характеризують рівень стану і зміни досліджуваного явища.

Аналіз часових рядів – сукупність математико-статистичних методів аналізу, призначених для виявлення структури часових рядів і для їх прогнозування. Сюди належать, зокрема, методи регресійного аналізу. Виявлення структури часового ряду необхідно для того, щоб побудувати математичну модель того явища, яке є джерелом аналізованого часового ряду. Прогноз майбутніх значень часового ряду використовується для ефективного прийняття рішень.

Питання для контролю знань

1. Поняття нечітких часових рядів
2. Методи моделювання часових рядів
3. Методи аналізу поведінки часових рядів.
4. Методи прогнозування поведінки часових рядів.
5. Моделювання часового ряду.

Лекція 4. Алгоритми Data Mining: кластеризація

Постановка задачі кластеризації та представлення результатів. Види кластерів. Міри близькості, засновані на відстаннях. Базові алгоритми кластеризації. Адаптивні методи кластеризації.

Ключові поняття

Кластер – це група об'єктів зі спільними властивостями.

Кластеризація – описова процедура, яка надає можливість провести розвідницький аналіз та вивчити структуру даних, але без жодних статистичних висновків.

Кластерний аналіз – процедура розбиття множини об'єктів на визначене чи невідоме число класів (кластерів) згідно з певним критерієм якості класифікації

Міри близькості - метрика, що характеризує ступінь подібності (схожості) j -ї та k -ї одиниць сукупності. Такою метрикою може бути відстань між ними S або коефіцієнт подібності R_{jk} . Близькі, схожі за вибраними метриками одиниці вважаються належними до одного типу, однорідними. Вибір

метрики є вузловим моментом кластерного аналізу, від якого залежить кінцевий варіант поділу сукупності на класи.

Питання для контролю знань

1. Постановка задачі кластеризації та представлення результатів.
2. Види кластерів.
3. Міри близькості, засновані на відстаннях.
4. Базові алгоритми кластеризації.
5. Адаптивні методи кластеризації.

Лекція 5. Дерева рішень як метод дейтамайнінгу

Сутність дерев рішень, особливості їх використання та побудови. Дерева рішень для дослідження альтернатив. Сутність класифікації альтернативних варіантів та методи її проведення. Загальний опис дерев класифікації. Способи побудови класифікаційних дерев рішень.

Ключові поняття

Дерево рішень – це схема (граф), яка відображає структуру задачі багатокрокового процесу прийняття рішень у вибраній сфері аналізу (класифікація набору даних або зразків, пошук оптимального рішення на множині альтернатив, структуризація проблеми, схема отримання логічного висновку за допомогою аналізу евристики (база правил)).

Питання для контролю знань

1. Сутність дерев рішень, особливості їх використання та побудови
2. Дерева рішень для дослідження альтернатив.
3. Сутність класифікації альтернативних варіантів та методи її проведення.
4. Загальний опис дерев класифікації
5. Способи побудови класифікаційних дерев рішень.

Лекція 6. Технології нейронних мереж та генетичні алгоритми

Визначення та еволюція нейронних мереж. Математична модель штучного нейрона. Програмне забезпечення нейромереж. Апаратне забезпечення нейромереж. Діапазон застосування нейромереж, їх переваги та недоліки.

Ключові поняття

Нейромережі – це надзвичайно спрощені моделі нервової системи людини, що можуть імітувати такі здатності людини, як навчання, узагальнення й абстрагування. Ці здатності дають моделям можливість імітувати подібні людині поведінку та режими і використовувати їх для прийняття рішень.

Генетичні алгоритми – це спроби слідувати еволюційним процесам біологічних систем, в яких відбувається процес розмноження і навчання.

Штучний нейрон – це деякою мірою є певним аналогом біологічного (компонента мозку, яка забезпечує можливість оброблення інформації), що допомагає відтворити процес мислення людського мозку через передачу електричних (електрохімічних) імпульсів між нейронами.

Питання для контролю знань

1. Визначення та еволюція нейронних мереж.
2. Математична модель штучного нейрона.
3. Програмне забезпечення нейромереж.
4. Апаратне забезпечення нейромереж.
5. Діапазон застосування нейромереж, їх переваги та недоліки.

Лекція 7. Сховища даних та оперативний аналіз даних (OLAP)

Визначення сховища даних, порівняння з базами даних, використання. Архітектура сховища даних. ETL-процеси (добування, перетворення й завантаження даних). Вітрини даних, куби даних, багатомірна модель даних. Архітектура OLAP-систем: MOLAP, ROLAP, HOLAP.

Ключові поняття

База даних (БД) — це організована структура, яка призначена для зберігання, зміни та обробки взаємозалежної інформації, переважно великих обсягів.

Сховище даних (англ. *data warehouse*) — предметно орієнтований, інтегрований, незмінний набір даних, що підтримує хронологію і здатний бути комплексним джерелом достовірної інформації для оперативного аналізу та прийняття рішень. В основі концепції сховища даних (СД) лежить розподіл інформації, що використовують в системах оперативної обробки даних (OLTP) і в системах підтримки прийняття рішень (СППР).

OLTP (англ. *Online Transaction Processing*) — онлайнова обробка транзакцій. Спосіб організації БД, при якому система працює з невеликими за розмірами транзакціями, що йдуть великим потоком, і при цьому клієнту потрібний від системи максимально швидкий час відповіді.

Термін OLTP застосовують також до систем (застосунків). OLTP-системи призначені для введення, структурованого зберігання і обробки інформації (операцій, документів) в режимі реального часу.

Системи підтримки прийняття рішень (СППР) — інформаційні системи, які використовують обладнання, програмне забезпечення, дані, базу моделей і роботу менеджера з метою підтримки всіх стадій прийняття рішень у процесі аналітичного моделювання.

OLAP (англ. *online analytical processing*, аналітична обробка у реальному часі) — це інтерактивна система що дозволяє переглядати різні підсумки по багатовимірних даних. Термін "в реальному часі" (англ. *online*) означає що нові результати отримуються протягом секунд, без довгого очікування на результат запити.

Вітрина (кіоск) даних (англ. *Data mart*) — це структура або шаблон доступу, який визначає правила отримання даних із сховища даних для певних груп користувачів або цілей. Вітрина даних є підмножиною сховища даних, масивом тематичної, спеціалізованої інформації яка орієнтована на конкретну галузь бізнесу або команду. На відміну від сховищ даних, які мають доступ до всіх даних підприємства, інформація в вітринах даних обмежена даними одного підрозділу.

Куб даних – процедура представлення даних (таких як якась подія, міра, факт) у розрізі певних вимірів або набору характеристик. Наприклад, у випадку OLAP, куб можна сформувати як події продажу певного товару у певній дочірній компанії в певний час. У часових рядах супутникових знімків характеристиками, що розглядаються, будуть широта, довгота, час; фактом буде піксель для даної координати простору / часу, який приймає супутник.

Багатомірна модель даних— база даних, що підтримує багатовимірну модель даних на концептуальному рівні та призначена для інтерактивного аналітичного опрацювання агрегованих історичних і прогнозованих даних.

ETL процеси (Extract, Transform, Load - витяг, перетворення та завантаження) — процес, який використовується в базах даних та, особливо, у сховищах даних та у засобах Business Intelligence для забезпечення їх роботи для підтримки прийняття рішень.

Питання для контролю знань

1. Сутність поняття сховища даних
2. Використання даних сховищ
3. Архітектура сховища даних
4. ETL-процеси (добування, перетворення й завантаження даних).
5. Вітрини даних, куби даних, багатомірна модель даних
- 6
7. Основні властивості пакету аналізу R.
8. Статистична обробка даних в R.

А
р
х
і
т
е
к
т
у

Плани практичних занять

Виконання кожного практичного заняття передбачає врахування засвоєння положень дисципліни і особливостей теми дослідження здобувача. Виконану роботу здобувач повинен захистити, а оцінка враховується при підсумковій атестації

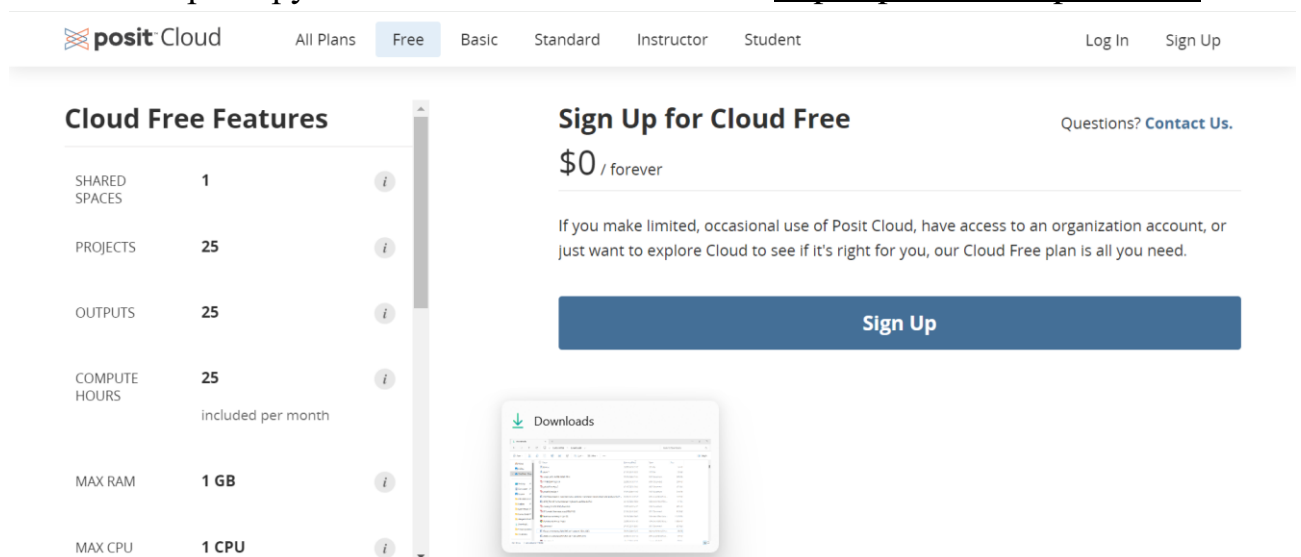
Робота здійснюється з використанням програмного забезпечення для обробки статистичних даних R Studio (комп'ютерна або хмарна версія) та MS Excel.

R — це мова програмування, яка використовується для статистичного аналізу, візуалізації та іншого аналізу даних.

RStudio Cloud (Posit Cloud) дає змогу користуватися хмарною версією програми. Для цього потрібний стабільний доступ до Інтернету.

Щоб отримати доступ до хмари RStudio, виконайте такі дії:

1. Зареєструйте обліковий запис на сайті <https://posit.cloud/plans/free>



The screenshot shows the Posit Cloud website interface. At the top, there is a navigation bar with the Posit Cloud logo and several tabs: 'All Plans', 'Free', 'Basic', 'Standard', 'Instructor', and 'Student'. The 'Free' tab is selected. On the right side of the navigation bar, there are links for 'Log In' and 'Sign Up'. Below the navigation bar, the main content area is divided into two sections. On the left, there is a table titled 'Cloud Free Features' with the following data:

Feature	Value	Info
SHARED SPACES	1	i
PROJECTS	25	i
OUTPUTS	25	i
COMPUTE HOURS	25	i
included per month		
MAX RAM	1 GB	i
MAX CPU	1 CPU	i

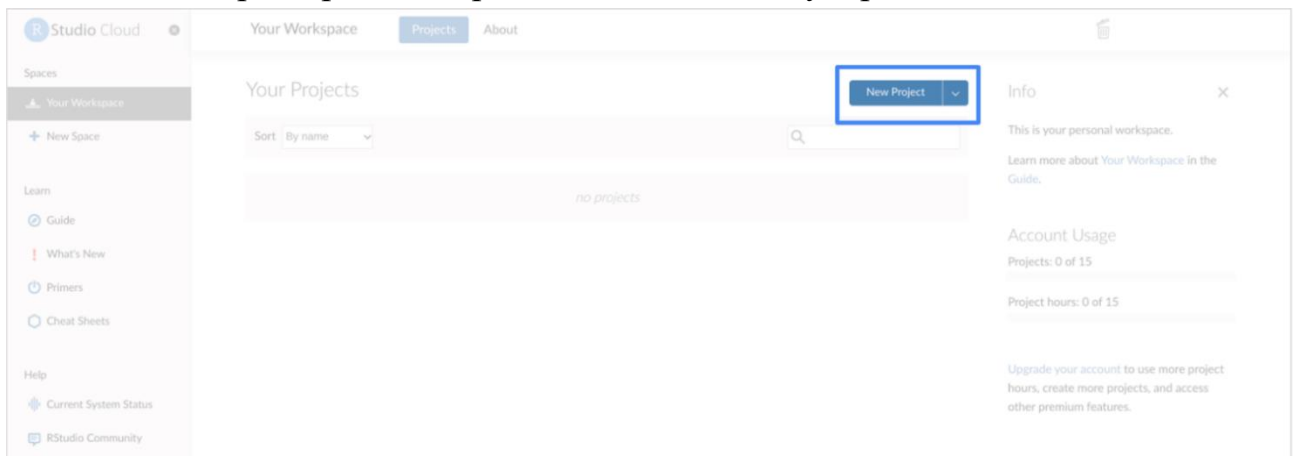
On the right, there is a section titled 'Sign Up for Cloud Free' with a sub-headline '\$0 / forever'. Below this, there is a paragraph: 'If you make limited, occasional use of Posit Cloud, have access to an organization account, or just want to explore Cloud to see if it's right for you, our Cloud Free plan is all you need.' At the bottom of this section is a large blue button labeled 'Sign Up'. To the right of the 'Sign Up' button, there is a link: 'Questions? Contact Us.' Below the 'Sign Up' button, there is a small inset image showing a 'Downloads' window with a list of files.

Тут знаходиться більше інформації про RStudio Cloud, включаючи тарифні плани. Протягом курсу ви використовуватимете безкоштовну версію, але вона має деякі обмеження – можна мати лише до 15 проектів у своєму безкоштовному обліковому записі та використовувати лише 15 проектних годин на місяць.

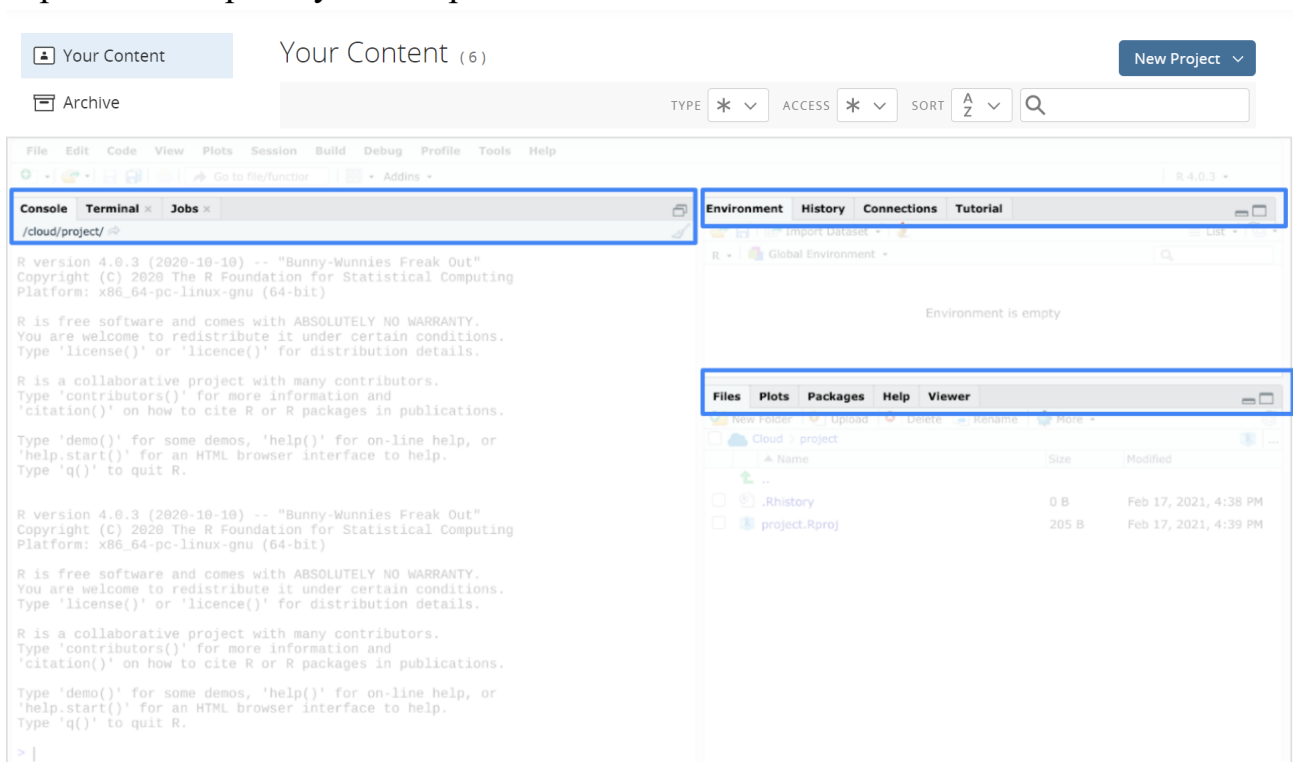
2. Натисніть кнопку «Зареєструватися» внизу праворуч, щоб розпочати користуватися безкоштовною версією.

Already have an account?
Log In
Sign Up

3. Введіть адресу електронної пошти, пароль, а також ім'я та прізвище.
4. Після реєстрації відкрийте RStudio Cloud уперше.



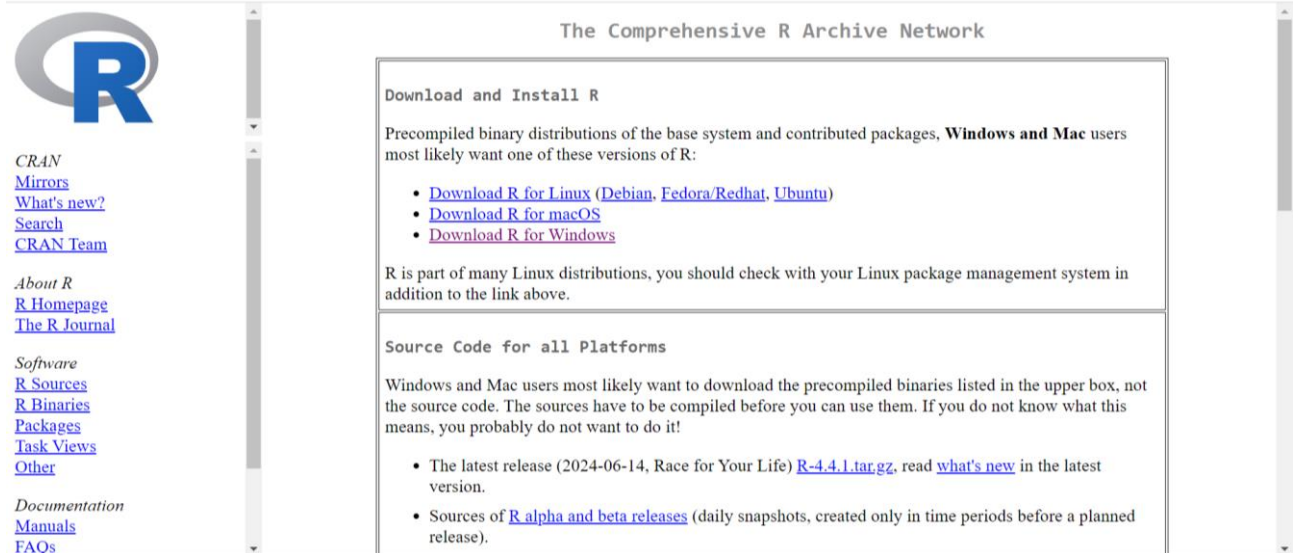
5. Клацніть «Новий проект» (New Project), щоб створити нове робоче середовище проекту та відкрити консоль RStudio Cloud.



Ви також можете встановити версію для настільного комп'ютера, яку можна завантажити відповідно до інструкцій. Це хороша альтернатива, якщо ви хочете мати можливість працювати з R офлайн.

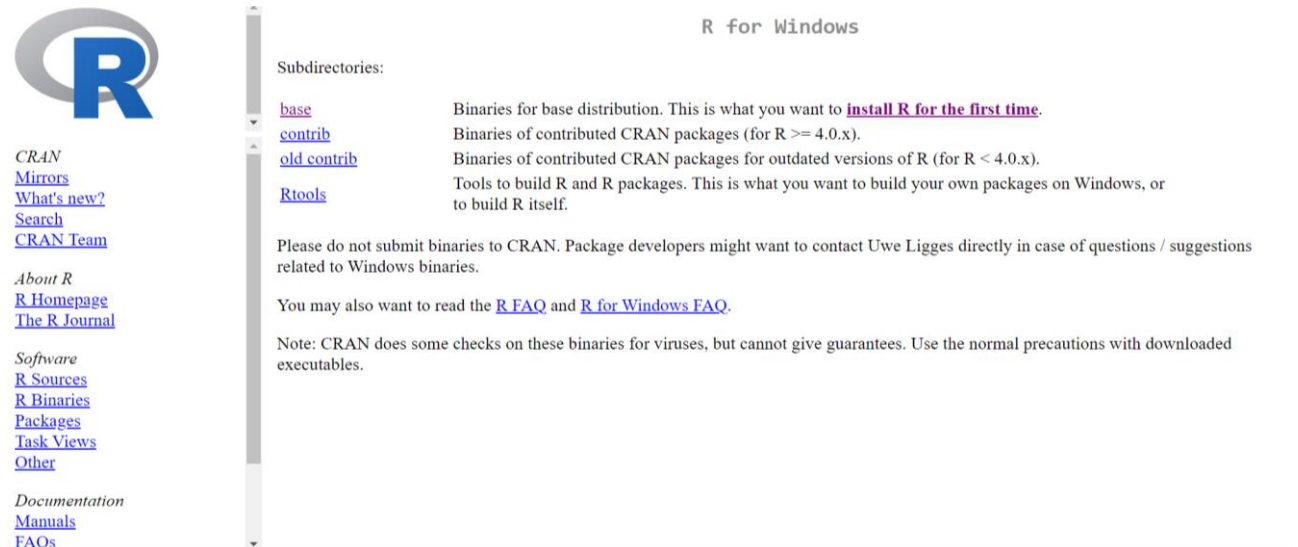
Встановлення R:

1. Перейдіть за посиланням <https://cloud.r-project.org/>
2. Оберіть вашу операційну систему



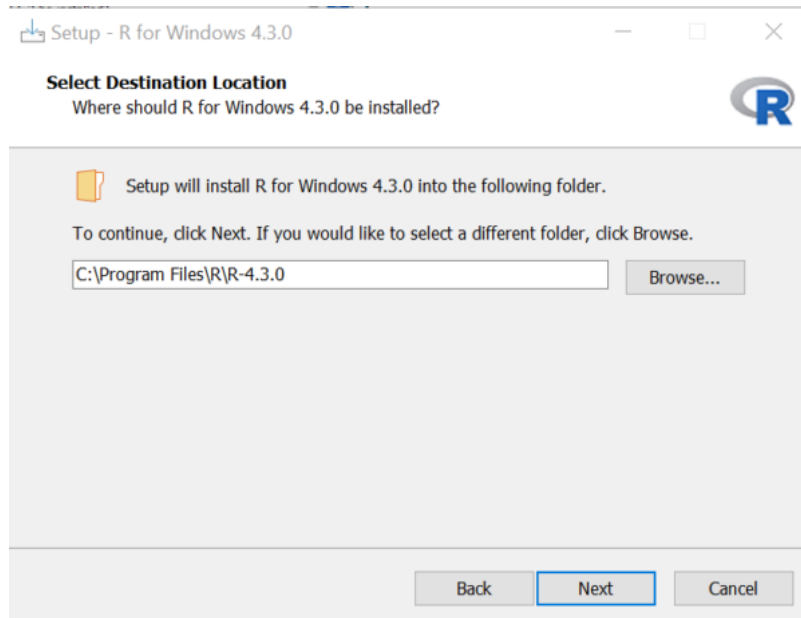
The screenshot shows the CRAN website interface. On the left is a navigation menu with links for CRAN, Mirrors, What's new?, Search, CRAN Team, About R, R Homepage, The R Journal, Software, R Sources, R Binaries, Packages, Task Views, Other, Documentation, Manuals, and FAQs. The main content area is titled 'The Comprehensive R Archive Network' and contains a section 'Download and Install R'. This section explains that precompiled binary distributions are available for Windows and Mac users. It lists three download options: 'Download R for Linux (Debian, Fedora/Redhat, Ubuntu)', 'Download R for macOS', and 'Download R for Windows'. Below this, it notes that R is part of many Linux distributions and suggests checking with the Linux package management system. A second section, 'Source Code for all Platforms', explains that Windows and Mac users should download precompiled binaries rather than source code. It lists two options: the latest release (R-4.4.1.tar.gz) and sources of R alpha and beta releases.

3. Завантажте відповідний пакунок



The screenshot shows the 'R for Windows' section of the CRAN website. The left navigation menu is identical to the previous screenshot. The main content area is titled 'R for Windows' and lists subdirectories: 'base', 'contrib', 'old contrib', and 'Rtools'. Each subdirectory has a brief description: 'base' (Binaries for base distribution), 'contrib' (Binaries of contributed CRAN packages for R >= 4.0.x), 'old contrib' (Binaries of contributed CRAN packages for outdated versions of R), and 'Rtools' (Tools to build R and R packages). A note at the bottom states: 'Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries. You may also want to read the R FAQ and R for Windows FAQ. Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.'

4. Інсталюйте його.



Далі слідуйте інформації у вашому комп'ютері для завантаження файлу. Після встановлення програми R ви можете перейти до налаштування RStudio.

Встановлення RStudio:

1. Перейдіть за посиланням <https://posit.co/download/rstudio-desktop/>

Grow your data science skills at [posit::conf\(2024\)](#)
August 12th-14th in Seattle [LEARN MORE](#) ✕

[posit](#) PRODUCTS ▾ SOLUTIONS ▾ LEARN & SUPPORT ▾ EXPLORE MORE ▾ PRICING Q

1: Install R

RStudio requires R 3.6.0+. Choose a version of R that matches your computer's operating system.

[DOWNLOAD AND INSTALL R](#)

2: Install RStudio

[DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS](#)

Size: 262.79 MB | [SHA-256: 09E1E38A](#) | Version: 2024.04.2+764 | Released: 2024-06-10

2. Якщо Ви працюєте з іншою операційною системою, то опустіться нижче на сторінці та перейдіть за покликанням, яке актуальне для вашої операційної системи.



Operating System	Download Link	File Size	Download ID
macOS 12+	RSTUDIO-2024.04.2-764.DMG	664.40 MB	D0DDD395
Ubuntu 20/Debian 11	RSTUDIO-2024.04.2-764-AMD64.DEB	194.73 MB	87B20155
Ubuntu 22/Debian 12	RSTUDIO-2024.04.2-764-AMD64.DEB	196.64 MB	1D0BD2F5

- Завантажте відповідний пакунок.
- Інсталюйте його.

RStudio — це інтерфейс (GUI) для полегшення використання R. Ви можете розглядати R як двигун транспортного засобу, який виконує важливу роботу, а RStudio — як кузов транспортного засобу (із сидіннями, аксесуарами тощо), який допомагає ви насправді використовуєте двигун, щоб рухатися вперед.

Типово RStudio відображає чотири прямокутні панелі. Якщо ваш RStudio відображає лише одну ліву панель, це тому, що у вас ще немає відкритих сценаріїв.

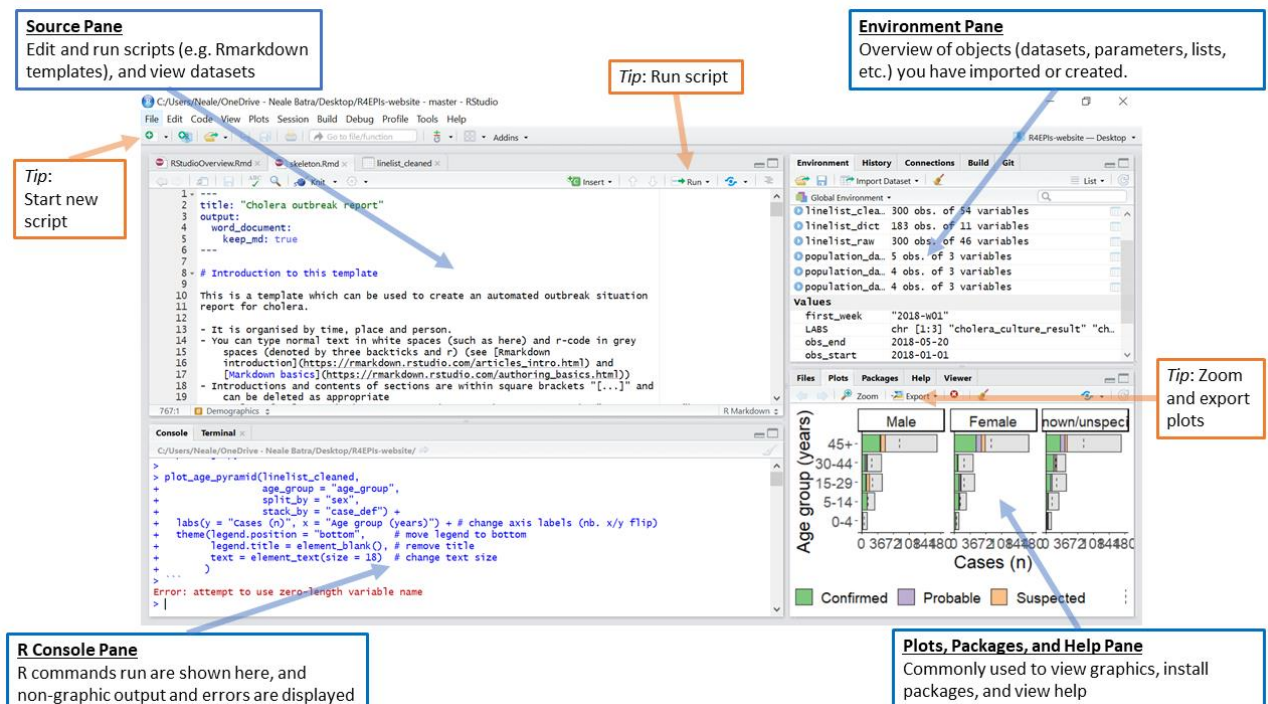


Рис. 1.1 – Серовище розробки R-Studio

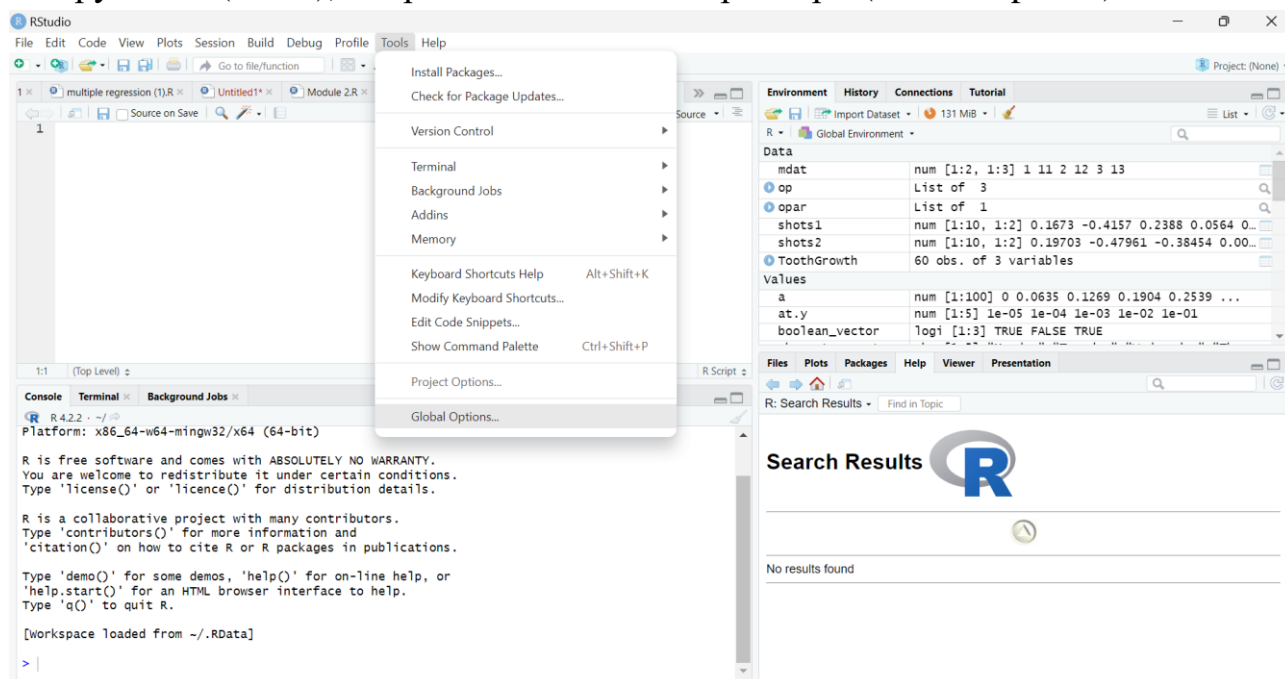
Панель консолі R (R Console Pane) – за замовчуванням ліва або нижня ліва панель у R Studio, є домом для «двигуна» R. Тут фактично виконуються команди та з'являються неграфічна інформація та повідомлення про помилки/попередження. Ви можете безпосередньо вводити та запускати команди в консолі R, але пам'ятайте, що ці команди не зберігаються в такому вигляді, як вони є під час виконання команд із сценарію.

Панель джерела (Source Pane) – за замовчуванням у верхньому лівому куті, є місцем для редагування та запуску ваших сценаріїв. Ця панель також може відображати набори даних (кадри даних) для перегляду.

Панель середовища (Environment Pane) – за замовчуванням верхня права, найчастіше використовується для перегляду коротких підсумків об'єктів у середовищі R у поточному сеансі. Ці об'єкти можуть включати імпортовані, змінені або створені набори даних, параметри, які ви визначили або вектори чи списки, які ви визначили під час аналізу (наприклад, назви регіонів). Клацніть стрілку поруч із назвою кадру даних, щоб побачити його змінні.

Діаграми, пакети та панель довідки (Plots, Packages and Help Pane) – нижня права панель містить кілька вкладок, включаючи графіки (відображення графіки, включаючи карти), довідку, бібліотеку файлів і доступні пакети R (включаючи параметри встановлення/оновлення).

Зміна налаштувань параметрів RStudio та вигляд можна у спадному меню Інструменти (Tools), вибравши Глобальні параметри (Global Options).



Усе, що ви зберігаєте в R — набори даних, змінні, наприклад список назв сіл, загальна чисельність населення, навіть такі результати, як графіки – це об'єкти, яким присвоюється ім'я і на які можна посилатися в наступних командах.

Об'єкт існує, якщо ви присвоїли йому значення. Коли йому присвоюється значення, об'єкт з'являється в середовищі (див. верхню праву панель RStudio). Тоді ним можна оперувати, маніпулювати, змінювати та перевизначати.

Символ «#» означає початок коментарю. Усе, що перебуває після цього знаку в одному рядку, ігнорується програмою.

Також можливе створення об'єктів за допомогою оператора присвоєння (<-). Наприклад: `object_name <- значення` (або процес/обчислення, які створюють значення)

Під час імпорту наборам даних також призначаються імена та визначаються як об'єкти.

Наприклад:

`linelist_raw` створюється та призначається значення імпортованого файлу CSV

```
linelist_raw <- rio::import(here("linelist.csv"))
```

`linelist_cleaned` створюється та присвоюється значення `linelist_raw`

```
linelist_cleaned <- linelist_raw
```

`linelist_cleaned` повторно визначений як сам, але змінений, щоб включити нову змінну

```
linelist_cleaned <- mutate(linelist_cleaned, obs_days =  
as.numeric(date_of_exit - date_of_admission))
```

У RStudio введення Alt + -(натисніть Alt, клавіша поруч із пробілом одночасно з -клавішею) буде писати <-одним натисканням клавіші.

Консоль RStudio надає цілий ряд опцій, що полегшують роботу з мовою R. Пропонуємо ознайомитися з ними нижче. Наприклад, автоматичне завершення коду: набираючи початок команди середовище пропонує користувачу продовження (рис.1.2).

Щоб виконати код, виділіть рядки та натисніть піктограму Run з зеленою стрілкою або комбінацію клавіш CTRL + ENTER.

Система R підтримує два режими роботи – інтерактивний режим і режим скрипта.

При запуску програми RGui автоматично з'являється вікно R Console. Це командне вікно (консоль), у якому користувач вводить команди, а програма друкує результати. У ході роботи в основному графічному вікні можуть з'явитися й інші вікна – редактор скриптів, вікна із графічним результатом виконання команд (графіки) тощо.

Команди вводяться користувачем у консолі (командному вікні) після символу запрошення, що має вигляд «>».

Після натискання кнопки Enter введена команда надходить на обробку. В одному рядку можна ввести кілька команд, розділяючи їх символом «;». Одну

команду можна розташувати на двох і більше рядках. Для цього слід натиснути Enter, тоді на новому рядку замість «>» з'явиться запрошення «+».

Кнопки стрілок «вгору» та «вниз» на клавіатурі дозволяють здійснювати навігацію серед раніше введених команд (можна вибрати одну з попередніх команд).

За допомогою кнопок стрілок «вліво» та «вправо» можна переміщатися у вже введеній команді, зокрема, редагуючи її.

R має модульну структуру. Ви встановлюєте базовий функціонал і розширяєте його потрібними вам бібліотеками. Зараз у репозиторії CRAN налічується ~ 7000 бібліотек. Для встановлення бібліотеки використовується команда `install.packages`.

Скопіюйте у R файл ці команди:

```
install.packages('dplyr', dependencies = TRUE)
```

```
install.packages('ggplot2', dependencies = TRUE)
```

Робота зі скриптами. Скрипт – це самостійно написана програма з використанням всіх можливостей R. Щоб створити новий скрипт, потрібно у командному меню вибрати Файл, а потім вибрати Новий скрипт. Відкриється нове вікно, що представляє із себе редактор, у якому можна писати та редагувати текст створюваної програми.

Запуск на виконання рядка або виділеного блоку – комбінація клавіш Ctrl+R.

R як калькулятор:

```
2+3-8 # додавання та віднімання
## [1] -3

7*5/2 # множення та ділення
## [1] 17.5

pi # константа pi
## [1] 3.141593

sqrt(4) # корінь квадратний
## [1] 2

2^3 # піднесення до степеня
## [1] 8
```

Символ присвоєння: <-

```
# x присвоїти значення 2
x <- 2

# вивести значення x
x
## [1] 2
```

x та X - різні змінні

```
X <- 3

X
## [1] 3

x
## [1] 2
```

Деякі основні команди R.

help(ім'я_функції) або *?ім'я_функції* – виклик контекстної допомоги.

ls() або *objects()* виводять на екран всі створені протягом сесії об'єкти (дуже корисно, якщо програма велика).

rm(ім'я_об'єкта) – видаляє об'єкт із зазначеним ім'ям.

example(ім'я_функції) – виводить приклад (якщо є) для обраної функції.

history(n) – буде виведене нове вікно, у якому перераховані n останніх команд.

getwd() – вивід поточної (робочої) директорії.

setwd('ім'я_нової_робочої_директорії') – зміна робочої директорії.

dir() – виводить список файлів у робочій директорії.

source('ім'я_файлу.R') – збереження R-коду в поточній директорії у файлі із зазначеним ім'ям і розширенням .R.

sink('ім'я_файлу.розширення') – перенаправляє потік виводу з екрану в зазначений файл; повторний виклик команди *sink()* закриває файл.

rnorm(n,mean,sd) – генерація послідовності *n* випадкових чисел, що мають нормальний розподіл з математичним очікуванням *mean* і середнім квадратичним відхиленням *sd*.

Типи даних в R.

Всі дані в R можна можуть мати наступні типи:

– *numeric* – об'єкти даного класу діляться на цілочислові (*integer*) і дійсні (*double*);

– *complex* – об'єкти комплексного типу;

– *logical* – логічні об'єкти, приймають тільки два значення: FALSE (F) і TRUE (T);

– *character* – символні об'єкти (символьні змінні задаються або в подвійних лапках, або в одинарних).

Дізнатися тип змінної можна з допомогою функції *class*.

```
v1 <- TRUE
class(v1)

## [1] "logical"

v1 <- -10.6
class(v1)

## [1] "numeric"

v1 <- 3L # L вказує що це ціле число
class(v1)

## [1] "integer"

v1 <- 3+2i
class(v1)

## [1] "complex"

v1 <- "stats"
class(v1)

## [1] "character"
```

Типи R об'єктів:

5. Вектор
6. Матриця
7. Список(*list*)
8. Фактор
9. Таблиця даних (*data frame*)

Вектор – набір значень одного типу. Утворюється з допомогою функції *c* (скорочення *concatenate*).

```
numeric_vector <- c(1, 10, 49)
```

```
boolean_vector <- c(TRUE, FALSE, TRUE)
```

```
character_vector <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday")
```

Операції з векторами:

```
x <- c(10, 2, 3, 7, 4)
```

```
y <- c(2, -1, 3, 2, 6)
```

```
# додавання/віднімання
```

```
# додаються/віднімаються поелементно
```

```
x + y
```

```
## [1] 12 1 6 9 10
```

```
x - y
```

```
## [1] 8 3 0 5 -2
```

```
# множення на скаляр
```

```
2*x
```

```
## [1] 20 4 6 14 8
```

```
# застосування функції до кожного елемента
```

```
sqrt(x)
```

```
## [1] 3.162278 1.414214 1.732051 2.645751 2.000000
```

```
# сума елементів
```

```
sum(x)
```

```
## [1] 26
```

```
# довжина вектора
```

```
length(x)
```

```
## [1] 5
```

```
# об'єднання векторів
```

```
z <- c(x, y)
```

```
z
```

```
## [1] 10 2 3 7 4 2 -1 3 2 6
```

Доступ до елементів вектора.

```
x <- 1:20 # інший спосіб задання вектора, вказуємо послідовність від 1 до 20
x
## [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
# п'ятий елемент(відлік починається з одиниці)
x[5]
## [1] 5
# елементи з 6 по 12
x[6:12]
## [1] 6 7 8 9 10 11 12
# елементи 6, 10, 13
x[c(6, 10, 13)]
## [1] 6 10 13
# елементи за винятком 6 та 13
x[-c(6, 13)]
## [1] 1 2 3 4 5 7 8 9 10 11 12 14 15 16 17 18 19 20
# елементи, які більше 5
x[x > 5]
```

```
## [1] 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
```

```
# елементи, які більше 5 і менші 15
x[x > 5 & x < 15]
```

```
## [1] 6 7 8 9 10 11 12 13 14
```

```
# елементи, які менші 5 або більші 15
x[x < 5 | x > 15]
```

```
## [1] 1 2 3 4 16 17 18 19 20
```

Відсутні значення (аналог null) позначаються як NA (Not available). Впливають на результат обчислень.

```
x <- c(10, 20, NA, 4, NA, 2)
```

```
sum(x)
```

```
## [1] NA
```

```
sum(x, na.rm = TRUE)
```

```
## [1] 36
```

Матриця – по суті двовимірний вектор.


```
mat <- matrix(data=c(9,2,3,4,5,6),ncol=3)
mat

##      [,1] [,2] [,3]
## [1,]    9    3    5
## [2,]    2    4    6
```

Список (list). Якщо елементи вектора мають бути одного типу, то елементи списку можуть мати різні типи.

```
a <- list(p_name="Joe", 4, foo=c(3,8,9))
print(a)

## $p_name
## [1] "Joe"
##
## [[2]]
## [1] 4
##
## $foo
## [1] 3 8 9
```

Доступ до елементів з використанням символу [[]] або \$ та імені(якщо елемент має ім'я)

Дата фрейм (data frame) використовується для роботи з таблицями.

Є три способи створити data frame.

1. Об'єднати вектори однакової довжини, використовуючи команду data.frame

```
cause <- c('pilot error', 'mechanical', 'weather', 'sabotage', 'other')
amount <- c(640, 195, 63, 95, 111)
plane_crash <- data.frame(cause, amount)
```

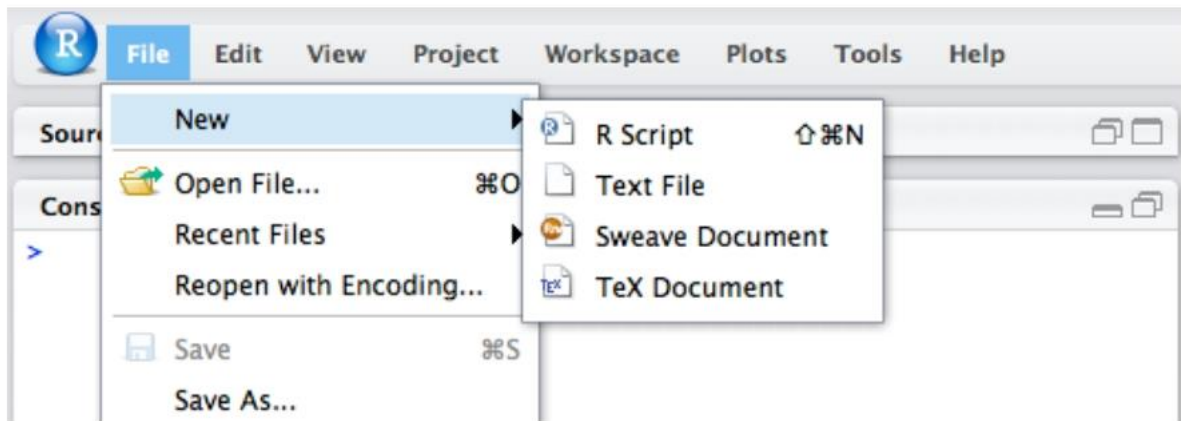
2. Використовуючи вбудовані набори даних

```
head(airquality)

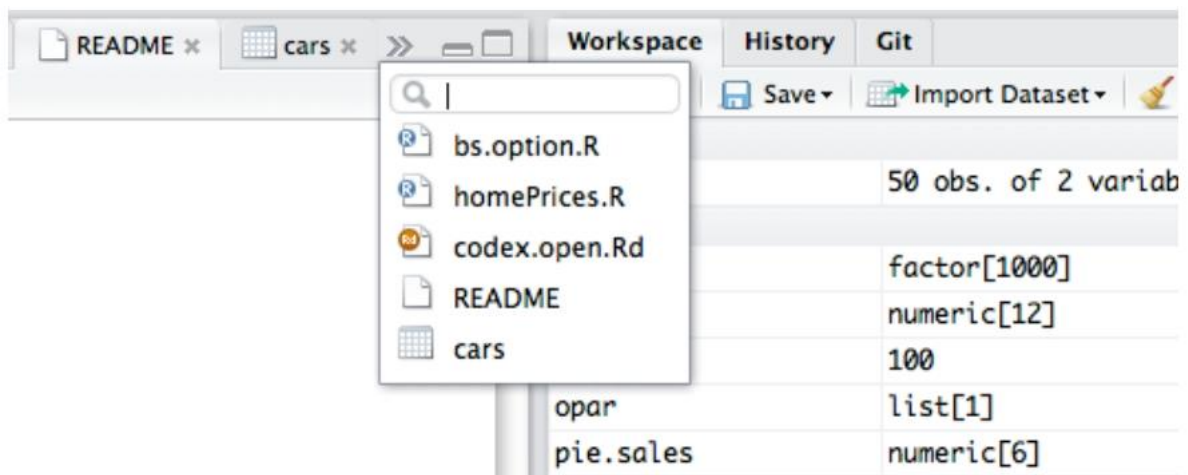
##      Ozone Solar.R Wind Temp Month Day
## 1     41     190   7.4   67     5    1
## 2     36     118   8.0   72     5    2
## 3     12     149  12.6   74     5    3
## 4     18     313  11.5   62     5    4
## 5     NA       NA  14.3   56     5    5
## 6     28       NA  14.9   66     5    6
```

3. Зчитати з файла

Для створення нового файлу використовуйте меню File -> New:

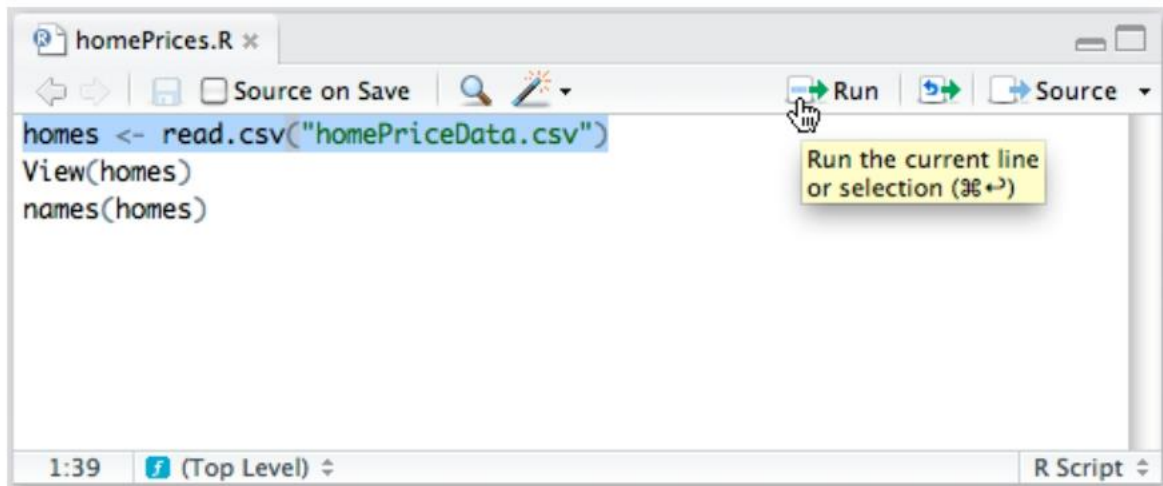


Для відкриття існуючого файлу необхідно скористатися меню File -> Open або Open Recent (для відкриття файлу, з яким робота здійснювалася недавно). Якщо відкрито кілька файлів одночасно, швидкий перехід від одного документа до іншого виконується за допомогою відповідних закладок у верхній частині вікна редактора коду. Для навігація між великою кількістю відкритих файлів служить іконка >> в правій верхній частині вікна Редактора; можна також скористатися меню View -> Switch to.



RStudio підтримує виконання коду безпосередньо з вікна Редактора (виконувані команди надсилаються в Консоль, де з'являється також результат їх виконання).

Для виконання поточного рядка коду можна скористатися поєднанням клавіш Ctrl + Enter або кнопкою Run Line (s), розташованої у верхній частині вікна Редактора:



Правилами підготовки завантаження даних з інших програм:

10. в імпортованій таблиці з даними не повинно бути порожніх клітинок. Якщо деякі значення з тих чи інших причин відсутні, замість них слід ввести NA.

11. імпортовану таблицю з даними рекомендується перетворити в простий текстовий файл з одним з допустимих розширень. На практиці зазвичай використовуються файли з розширенням *.txt*, в яких значення змінних розділені знаками табуляції (tab-delimited files), а також файли з розширенням *.csv* (comma separated values), в яких значення змінних розділені комами.

12. як першого рядка в імпортованій таблиці рекомендується ввести заголовки стовпців-змінних. Така рядок - зручний, але не обов'язковий елемент завантаження. Якщо вона відсутня, то про це необхідно повідомити в описі команди, яка буде керувати завантаженням файлу (наприклад, `read.table()` - див. Нижче). Усі наступні рядки файлу в якості першого елемента містять заголовки рядків (якщо такі передбачені), після яких слідує значення кожної з наявних в таблиці змінних. В іменах стовпців таблиці не допускається наявність прогалін. Крім того, імена стовпців (так само як і імена рядків) не повинні починатися з точки або чисел. Щоб уникнути пов'язаних з кодуванням проблем все текстові величини в імпортованих файлах рекомендується створювати з використанням літер латинського алфавіту.

13. файл який підлягає імпортуванню рекомендується помістити в робочу папку програми, тобто папку, в якій R за замовчуванням буде "намагатися знайти" цей файл. Щоб з'ясувати шлях до робочої папки R на своєму комп'ютері використовуйте команду `getwd()` (get working directory - дізнатися робочу директорію); наприклад:

```
getwd()
[1] "C:/Temp/"
```

Змінити робочу директорію можна за допомогою команди `setwd()` (set working directory - створити робочу директорію):

```
setwd("C:/My Documents")
```

Для завантаження підготовлених файлів досить використовувати мінімальний набір аргументів функції `read.table()`.

`read.table()`: функція `read.table` є найбільш зручним способом зчитування в прямокутній сітці даних.

`read.table("file1.txt", header = TRUE, row.names = 1, sep = ",")`: `header = TRUE`: Файл містить назви змінних як його перший рядок.

`row.names = 1`: це може бути вектор, що дає фактичні назви рядків, або одне число, що дає стовпець таблиці, що містить назви рядків (поточний сценарій) або рядок символів, що дає ім'я стовпця таблиці, що містить рядок імена. Якщо `row.names` відсутній, рядки автоматично пронумеруються.

`sep = ","`: Використовується для розділення між полями. Типовим параметром є пробіл.

Відсутні значення: за замовчуванням файл передбачає, що він містить рядок символів NA для представлення відсутніх значень, але це може бути змінено аргументом `na.strings`, який є вектором одного або декількох символічних зображень відсутніх значень.

Як приклад припустимо, що нам необхідно завантажити файл `hydro_chem.txt`, який зберігається в робочій папці R і містить дані за хімічним складом води деякого водойми. Завантажується таблицю даних ми маємо намір зберегти як об'єкта з ім'ям `chem`. Функції `read.table()` в цьому випадку може бути застосована в такий спосіб:

```
chem <- read.table(file = "hydro_chem.txt", header = TRUE)
```

Якщо у вас є доступ до Excel, експортуйте потрібні дані з Excel у форматі, відокремленій табуляцією або розділеними комами, і використовуйте `read.delim()` або `read.csv()`, щоб імпортувати його в R. `read.delim()` або `read.csv()` подібні до `read.table()`.

Крок 1: Збережіть файл excel як файл `fuel.csv`.

Крок 2: у R: `fuel <- read.csv("fuel.csv", header = T)`

Як зазначено вище, часто імпортовані в R файли мають формат `csv`. Для їх завантаження можна скористатися тією ж функцією `read.table()`, але при цьому слід вказати, що як роздільник значень змінних у файлі використовується кома:

```
chem <- read.table(file = "hydro_chem.csv", header = TRUE, sep = ",")
```

Аналогом `read.table()` для зчитування `csv`-файлів є функція `read.csv()`:

```
chem <- read.csv(file = "hydro_chem.csv", header = TRUE)
```

Якщо підлягає завантаженні файл зберігається в папці, відмінній від робочої папки R, то слід вказати повний шлях до нього. При цьому користувачам операційних систем Windows необхідно пам'ятати, що для вказівки повних шляхів до файлів в програмі R використовується не зворотний одинарний слеш (`\`), а прямий одинарний (`/`) або подвійний зворотний слеш (`\\`). Наприклад,

наступні дві команди будуть успішно сприйняті R і приведуть до ідентичному результату - завантаженні файлу hydro_chem.txt і збереженню його у вигляді об'єкта chem:

```
chem <- read.csv (file = "D: \\ Documents \\ hydrochem.txt", header = TRUE)  
chem <- read.csv (file = "D: /Documents/hydrochem.txt", header = TRUE)
```

Для інтерактивного вибору завантаження, який зберігається поза робочої папки R, можна застосувати допоміжну функцію file.choose () (вибрати файл). Виконання цієї команди призводить до відкриття звичайного діалогового вікна операційної системи Windows, в якому користувач вибирає папку з необхідним файлом. Дуже зручно поєднувати file.choose () з командами read.table () або read.csv (), наприклад:

```
chem <- read.table(file = file.choose(), header = TRUE, sep = ",")
```

Детальніше ознайомитися з алгоритмом роботи з векторами можна за покликанням <https://r4ds.had.co.nz/vectors.html#vectors>

Навчально-методичне забезпечення

1. Савчук В.К., Воляк Л. Р. Електронний курс «Дейтамайнінг» URL: <https://elearn.nubip.edu.ua/enrol/index.php?id=4412>

2. Нормативні документи.

. Програмно – методичне забезпечення навчально-наукової лабораторії біоеконометрики та дейтамайнінгу

М

о
д
и
ч
н
і

в
к
а
з
і
в
к
и

д
о

КОНТРОЛЬНІ ЗАХОДИ ПЕРЕВІРКИ ЗНАНЬ

(Приклад екзаменаційного білета)

I. Теоретичні питання (кількість балів за кожне розкрите теоретичне питання – 5 балів)

1.1. Розкрийте сутність концепції дейтамайнінгу та його застосування

1.2. Розкрийте сутність інтелектуального аналізу часових рядів

II. Тестові завдання (кількість балів за кожний тест – 2 бали)

Питання 1. До яких рядів динаміки належать показники, що характеризують розміри явища за певні проміжки часу?	
1	Моментні
2	Дискретні
3	Інтервальні
4	Ряди середніх
Питання 2. Для застосування кореляційного аналізу потрібні наявні передумови	
1	Достатня варіація досліджуваної ознаки
2	Числовий вираз ознаки
3	Достатня однорідність досліджуваної сукупності
4	Неоднорідність досліджуваної сукупності
Питання 3. Древа рішень відносяться до групи ...	
1	кібернетичних методів
2	логічних методів
3	статистичних методів
4	методів крос-табуляції
Питання 4. Одним з алгоритмів навчання нейронної мережі, які найчастіше використовуються є алгоритм...	
1	прямого розповсюдження помилки
2	зворотного розповсюдження помилки
3	диференційний
4	лінійний
Питання 5. Data Mining - це процес виявлення в сирих даних	
1	неочевидних закономірностей
2	раніше сформульованих гіпотез
3	великої кількості закономірностей
4	об'єктивних закономірностей
Питання 6. До статистичних методів Data Mining відносять:	
1	Аналіз зав'язків
2	Древа рішень;

3	Аналіз часових рядів
4	Кластерний аналіз
5	Штучні нейронні мережі
Питання 7. Вектор у середовищі R можна створити за допомогою функції:	
1	#
2	c
3	sum
4	length
5	<-
Питання 8. Для виведення даних описової статистики (сума, дисперсія тощо) у середовищі R використовується функція:	
1	Var()
2	Sd()
3	Summary()
4	View()
Питання 9. Для побудови лінійної регресійної моделі у середовищі R використовується функція:	
1	lm()
2	str()
3	plot()
4	fit()
5	predict()
Питання 10. Щоб змінити робочу директорію у середовищі R використовується функція:	
1	getwd ()
2	setwd ()
3	ls()
4	read.table ()

РЕКОМЕНДОВНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

1. Акіменко В.В., Загородній Ю.В. Проектування СППР на основі нечіткої логіки: Навч.- методич. посіб. Київ: КНУ, 2007. 94 с.
2. Барретт Л. Ф. 7 ½ уроків про мозок. Пер. з англ. Я. Лебеденка. Харків: КСД, 2021. 207 с.
3. Воляк Л. Р. Моделювання та кількісний вимір впливу основних факторів на продуктивність рослинництва. Науковий вісник Національного університету біоресурсів і природокористування України. Серія «Економіка, аграрний менеджмент, бізнес»/ редкол.: С. М. Ніколаєнко (відп. ред.) та ін. – К.: ВЦ НУБіП України, 2016. –Вип. 249,. –506 с., С. 117-125.
4. Гладун Ф.Я., Рогушина Ю.В. Data Mining: пошук знань в даних: підручник. Київ: ТОВ «ВД «АДЕФ-Україна», 2016. 452 с.
5. Гнатюк В. Вступ до R на прикладах. Харків: Харківський національний економічний університет, 2010. 101 с.
6. Гончар Л. І., Чирка Д. М. Застосування дейтамайнінгу в бізнесовій діяльності. URL: <https://core.ac.uk/download/pdf/78513614.pdf>
7. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. Харків: ХНУРЕ, 2021. 92 с.
8. Грінгарт Семюель. Інтернет речей. пер. з англ. Л. Герасимчука. Харків: КК, 2018. 175 с.
9. Гузь М. М., Музиченко А. О. Нові підходи до управління виробництвом сільського господарства: органічне землеробство та безпека харчових продуктів. Статистико-аналітичне забезпечення управління інноваційним розвитком економічних суб'єктів: кол. моногр./ за заг. ред В. К. Савчука. –К.: ФОП Ямчинський О., 2020. – 292 с., С.101 – 119.
10. Данильченко О. М., Данильченко А.О. Інтелектуальний аналіз даних: Навч. посібник. - Житомир: ЖДТУ, 2009. - 405 с.
11. Додонов О., Кузьмічов А. Дейтамайнінг в Excel. Розвідувальний аналіз даних. Ліра-Київ: 2023. 240 с.
12. Куць О. І., Куць Т. В. Аналітична оцінка фінансового забезпечення інноваційного розвитку підприємств аграрного сектору. Статистико-аналітичне забезпечення управління інноваційним розвитком економічних суб'єктів: кол. моногр./ за заг. ред В. К.Савчука. –К.: ФОП Ямчинський О., 2020. – 292 с., С.270 – 281.
13. Лазаришина І. Д. Основні теоретичні засади економічного аналізу додаткової вартості. Наукові праці Кіровоградського національного технічного університету. Економічні науки : зб. наук. пр. Кіровоград : КНТУ, 2006. Вип. 10, ч. 2. С. 135–139.
14. Майборода Р.Є. "Комп'ютерна статистика". ВПЦ "Київський університет", 2019. - 589 с. URL: <http://probability.univ.kiev.ua/userfiles/mre/cscolor.pdf>

15.Майборода Р.Є., Сугакова О.В. Аналіз даних за допомогою пакета R. Навчальний посібник. Київ, Київський національний університет ім. Тараса Шевченка, 2015. 65 с.

16.Макарчук О. Г. Деа-аналіз як метод прийняття рішень. Науковий вісник Національного університету біоресурсів і природокористування України. Серія «Економіка, аграрний менеджмент, бізнес»/ редкол.: Д. О. Мельничук (відп. ред.) та ін. – К.: ВЦ НУБіП України, 2012. –Вип. 169, ч. 2. –338 с., С. 190-195.

17.Манжосова І. Б. PESTLE-аналіз чинників, які впливають на впровадження цифрових технологій в сільському господарстві (рос.). Стратегічні пріоритети розвитку аграрних формувань: аналітико-прогнозні тренди: кол. моногр. За заг. ред. В. К. Савчука. – К.: ЦП «Компринт», 2018. – 375 с., С.6 – 22.

18.Савчук В. К. Методичні підходи і моделі прогнозування стратегічного розвитку підприємств. Стратегічний розвиток підприємств аграрної сфери економіки України: аналітико-прогнозна оцінка: кол. моногр./ за заг. ред. В. К. Савчука - К.: ЦП «Компринт», 2017. – 366 с., С. 353 – 360.

19.Савчук В. К., Гакі П. К. Мета-аналіз як спосіб підвищення доказовості результатів дослідження. Науковий вісник Національного університету біоресурсів і природокористування України. Серія «Економіка, аграрний менеджмент, бізнес»/ редкол.: С. М. Ніколаєнко (відп. ред.) та ін. – К.: ВЦ НУБіП України, 2018. –Вип. 290, –283 с., С. 219-228.

20.Симоненко О. І. Методичні підходи до економетричного моделювання стратегій інноваційного розвитку аграрних формувань. Статистико-аналітичне забезпечення управління інноваційним розвитком економічних суб'єктів: кол. моногр./ за заг. ред В. К.Савчука. –К.: ФОП Ямчинський О., 2020. – 292 с., С.231 – 252.

21.Ситник В.Ф., Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг): навч. посіб. Київ: КНЕУ, 2007. 376 с.

22.Черняк О.І., Захарченко П.В. Інтелектуальний аналіз даних: підручник. Київ: Знання, 2014. 599 с.

23.Чухліб А. В. Моделювання стратегічної конкурентоспроможності сільськогосподарських підприємств: теоретико-методичний аспект. Статистико-аналітичне забезпечення управління інноваційним розвитком економічних суб'єктів: кол. моногр./ за заг. ред В. К.Савчука. –К.: ФОП Ямчинський О., 2020. – 292 с., С.253 – 269.

24.Шпігельхальтер Д. Мистецтво статистики. Прийняття аргументованих рішень на основі даних. Київ. 2023. 384 с.

25.Bai, Ju. Special Issue on Big Data / Ju. Bai, Ji. Fan and R. Tsay // Journal of Business & Economic Statistics. 2016. vol. 34. issue 4. P. 487-488.

26.Big data: The next frontier for innovation, competition, and productivity. Report / James Manyika, Michel Chui, Brad Brown, Jacques Bughin, Richard Dobbs,

Charles Roxburgh, Angela Hung Byers. McKinsey Global Institute. May 2011. URL: <https://www.mckinsey.com/businessfunctions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation> .

27. Jiawei Han, Micheline Kamber, Jian Pei. Data Mining Concepts and Techniques. Third Edition. 2012. 740 p. URL: <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>

28. Manyika James, Chui Michel, Brown Brad, Bughin Jacques, Dobbs Richard, Roxburgh Charles, Byers Angela Hung. Big data: The next frontier for innovation, competition, and productivity. Report. McKinsey Global Institute. May 2011. URL: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation> .

29. Rinkesh Jain, Divakar Singh. Data Mining and Analysis of Economic Data. 2013. URL: https://www.researchgate.net/publication/337307254_Data_Mining_and_Analysis_of_Economic_Data

30. Yuliia V. Dehtiarova, Yuri Yevdokimov. Data Mining Methods and Models for Social and Economic Processes Forecasting. 2018. URL: https://mer.fem.sumdu.edu.ua/content/articles/issue_37/YULIIA_V_DEHTIAROVA_YURI_YEVDOKIMOVDData_Mining_Methods_and_Models_for_Social_and_Economic_Processes_Forecasting.pdf

31. Zumel N., Mount J. Practical Data Science with R. Manning Publications Co., 2014. 417 p. <https://www.manning.com/books/practical-data-science-with-r>

В

Електронні ресурси

1. Big Data for Development: From Information- to Knowledge Societies. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2205145
2. Food and Agriculture Organization of the United Nations. URL: https://commission.europa.eu/index_en
3. Official website of the European Commission. URL: https://commission.europa.eu/index_en
4. Official website of the United Nations. URL: <https://www.un.org/en/>
5. Official website of World Bank. URL: <http://www.worldbank.org/>
6. U.S. DEPARTMENT OF AGRICULTURE. URL: <https://www.usda.gov/>
7. Верховна Рада України. URL: <https://www.rada.gov.ua/>
8. Газета «Все про бухгалтерський облік». URL: <http://gazeta.vobu.ua/>
9. Головний сайт про агробізнес. URL: <https://latifundist.com/>
10. Державна служба статистики України. URL: <https://www.ukrstat.gov.ua/>

11. Електронна бібліотека НУБіП України URL:
<https://nubip.edu.ua/node/17325>.
12. Кабінет Міністрів України. URL: <https://www.kmu.gov.ua/>
13. Міністерство аграрної політики та продовольства України. URL:
<https://minagro.gov.ua/>
14. Міністерство Фінансів України. URL: <https://www.mof.gov.ua/uk>
15. Фахівець пояснив, чому парадокси важливі у бізнес-аналітиці.
URL: www.bigdatalab.com.ua › news