

ЗАСТОСУВАННЯ ОСНОВНИХ СПЕЦІАЛІЗОВАНИХ СТАТИСТИЧНИХ ПАКЕТІВ ДЛЯ ОБРОБКИ ІНФОРМАЦІЇ

Бойко Я.А.

IV курс, фізико-математичний факультет

Медведева Марія Олександрівна, канд.пед.наук, доцент

Уманський державний педагогічний університет імені Павла Тичини

Умань

Багато методів аналітики і моделювання, які використовувалися протягом багатьох років, зазнали істотних змін. Деякі з них, як лінійна регресія чи дерево ухвалення рішень, є все ще ефективними і релевантними, але досить спрощеними для використання. В попередній період існували певні обмеження стосовно наявності функціоналу та масштабованості. Це означало, що простота була однією з вимог для обробки даних.

Сучасні технології призвели до збільшення кількості даних. Для обробки таких великих масивів даних було створено сучасні аналітичні засоби, які дали змогу аналізувати дані точніше і детальніше. У зв'язку із зростанням кількості даних і необхідністю поглибленого аналізу деякі методи, як лінійна регресія чи дерево ухвалення рішень, були замінені на більш сучасні. До таких аналітичних методів відносимо: ансамблевий метод, моделювання продукту та аналіз текстових даних. Ансамблевий метод – дуже простий з концептуальної точки зору. Замість побудови єдиної моделі за допомогою одного підходу декілька моделей створюються за допомогою декількох засобів аналізу даних. Результати, які отримані з усіх моделей, об'єднуються для отримання остаточного рішення. Ефективність ансамблевого методу полягає в тому, що різні підходи презентують свої сильні та слабкі сторони. Поєднуючи розрахунки різних моделей за допомогою ансамблевого методу, кінцевий алгоритм показує кращі результати в цілому. Такий феномен часто називають «мудрістю натовпу».

До найбільш вживаних засобів аналізу даних належить проект R, SPSS та SAS [5]. Інструмент аналітики R використовується великою кількістю аналітичних фахівців. Це особливо актуально в академічних і дослідницьких

середовищах. R, як правило, застосовується для досліджень і розробок, а не у великомасштабних, критичних процесах виробництва. В корпоративному середовищі сьогодні, якщо є велика команда фахівців в сфері аналітики, часто трапляється, що принаймні кілька членів команди використовують R [4].

Аналітичний інструмент SPSS вперше був представлений в 1968 році. Його назва була змінена на SPSS Statistics в 2009 році після придбання SPSS-бізнесу IBM Group. Інструмент SPSS містить зручний GUI. Назва програмного забезпечення перекладається як статистичний пакет для соціальних наук (Statistical Package for the Social Sciences)(SPSS), що відображає початкову направленість і ринок продукту, хоча зараз він популярний і в інших галузях, включаючи медичні науки та маркетинг. Програма SPSS не містить всіх функціональних можливостей R, але його синтаксис і формат бази даних сумісні з R, і він може обробляти великі обсяги даних.

Система статистичного аналізу (SAS) була заснована в 1976 році в світі мейнфреймів IBM для обробки великих масивів даних. Його спроможність обробляти дані зросла з впровадженням паралельної архітектури в 1996 році [2].

Програма SAS складається з кроків даних (DATA steps), процедурних кроків (procedure steps) та макросів, якщо це необхідно. Процедури забезпечують повний спектр функцій (статистику, графіку, утиліти тощо), а крок даних дозволяє користувачеві відкривати файли (або імпортувати бази даних), зчитувати кожен запис по черзі, записувати в інший файл (або експортувати його до бази даних), об'єднувати декілька файлів і закривати файли [3, с. 36].

Отже, аналітичні інструменти R, SAS і SPSS відрізняються один від одного можливостями і займають різні ринкові ніші, а тому їх необхідно використовувати в залежності від поставленої мети, рівня кваліфікації аналітиків і фінансових можливостей підприємств.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бідюк, П.І. Прикладна статистика / П.І. Бідюк, О.М. Терентьєв, Т.І. Просянкіна-Жарова. – Вінниця: ПП «Едельвейс і К», 2013. – 288 с.

2. Приступая к программированию в SAS Studio 3.2 [Electronic resource]. – URL
https://support.sas.com/documentation/cdl_alternate/ru/webeditorgs/67431/PDF/default/webeditorgs.pdf
3. Терентьев А.Н. SAS BASE: Основы программирования / Терентьев А.Н., Домрачев В.Н., Костецкий Р.И. – К: Эдельвейс, 2014. – 304 с.
4. Castanedo Federico. New scalable solutions for data analysis with R. Comparison [Electronic resource]. – URL :
<https://www.oreilly.com/ideas/new-scalable-solutions-for-data-analysis-with-r>
5. R vs SAS vs SPSS – Top 3 Data Analytics tools Comparison [Electronic resource]. – URL : <https://data-flair.training/blogs/r-sas-spss-data-analytics-tools-comparison/>