

УДК 620.91+ 621.31

Іщенко О.С., аспірантка
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Україна

ЗАСТОСУВАННЯ ЗАСОБІВ R ДЛЯ ПРОГНОЗУВАННЯ ЦІН НА ЕЛЕКТРИЧНУ ЕНЕРГІЮ

Анотація. При створенні моделей прогнозування існують суперечки пов'язані з особливостями баз даних. Це призводить до складності прийняття рішень, що обумовлено різними чинниками які впливають на збір та оброблення інформації. Існує багато інструментів для підвищення якості обробки баз даних, і наступною проблемою є пошук оптимального вибору одного з запропонованих інструментів, який би максимально задовольнив усі нагальні потреби для побудови моделі прогнозу цін на ринках електричної енергії. В роботі проаналізовані основні переваги мови програмування R, розглянуті можливості сфери застосування, проведений порівняльний аналіз з іншими мовами програмування. R Studio є інтегрованим середовищем розробки і дозволяє користувачам розробляти та редагувати програми на R, підтримуючи велику кількість статистичних пакетів, більш високу якість графіки та можливість керувати робочим простором. Зручність інтерфейсу, можливість конвертування коду для подальшого його використання у іншій мові програмування, а також безкоштовність використання R Studio робить це інтегроване середовище розробки унікальним варіантом для побудови моделі прогнозування ціни на ринку електричної енергії України.

Ключові слова: прогнозування цін, ринок електроенергії, R, R Studio, Python, обробка баз даних.

Abstract. Creating forecasting models, there are disputes related to different database features. This leads to the complexity of decision making, which is due to various factors that affect the collection and processing of information. There are many tools for improving the quality of information processing, and the next problem is to find the optimal choice of one of the proposed tools that would maximally satisfy all urgent needs to build a model of pricing in the electricity markets. The main advantages of R programming language are analyzed in the paper, the scope of application is considered, comparative analysis with other programming languages is made. R Studio is an integrated development environment that allows users to develop and edit applications on R, supporting a large number of statistical packages, higher quality graphics and the ability to manage the workspace. The convenience of the interface, the ability to convert the code for further use in another programming language, as well as the free use of R Studio make this integrated development environment a unique option for building a model for predicting the price of the electricity market in Ukraine.

Key words: price forecasting, electricity market, R, R Studio, Python, database processing.

Вступ.

При виборі методів і побудові моделей прогнозування ціни на електричну енергію виникають складності, ті схеми і принципи побудови які припустимі для звичайних товарів не можуть бути застосовані в умовах обмежень ринку електроенергії [1-4]. Якісний аналіз даних при обмежених умовах є першочерговим завданням при моделюванні.

Аналіз даних - складова частина будь-якого дослідження, так як саме на етапі аналізу даних відбувається перехід від отриманих результатів до розуміння наявної в них інформації і далі до нового наукового знання. Практично жодне дослідження в енергетиці не може бути розглянуто якщо воно не має наукову доказовість, якщо висновки за підсумками проведеного

дослідження не підтверджені результатами статистичної обробки даних [5]. Широке впровадження комп'ютерних програм дозволило істотно прискорити процес аналізу даних, а також виконати ряд інших функцій. Зокрема, спеціалізоване програмне забезпечення допомагає підготувати дані для аналізу і здійснити його.

На практиці найбільш часто для статистичного аналізу даних використовуються такі програми, як Python, SAS, SPSS та ін. Повноцінною безкоштовною альтернативою цим комерційним продуктам є безкоштовне програмне середовище R, яке по суті є мовою програмування. Створена більше 20 років тому, ця мова успішно розвивається і підтримується науковим співтовариством у всіх областях наукового пошуку. Мова R добре документована, і для навчання користувачів існують інтернет-ресурси [6-10]. Про зростання популярності мови R свідчить той факт, що в рейтингу мов програмування TIOBE, який складається на основі частоти пошукових запитів, на травень 2020 року мова R знаходиться на 10-му місці, причому рейтинг виріс за рік на 11 позицій [11]. У цьому рейтингу з числа мов програмування, зазвичай використовуваних при аналізі даних, R випереджає тільки Python, але цей факт пояснюється тим, що сфера застосування Python істотно ширше сфери застосування R. В даній роботі розглянуто систему статистичного аналізу R та її інтегроване середовище розробки R Studio як найоптимальнішу з існуючих мов програмування саме для прогнозування цін на електричну енергію в умовах обмежень притаманних ринку електричної енергії України.

Мета та завдання.

Метою роботи є визначення можливостей мови програмування R та прикладних програм на її основі для моделей та методів прогнозування цін на електричну енергію.

Для досягнення поставленої мети були визначені наступні завдання.

I Огляд мови програмування R та прикладних програм на її основі.

II Порівняльний аналіз R з альтернативними мовами програмування та програмним забезпеченням.

III Визначення можливостей застосування R для моделей та методів прогнозування цін на електричну енергію.

Матеріал і результати досліджень.

I Огляд мови програмування R та прикладної програми на її основі

R - система статистичного аналізу, створена Россом Іхакою та Робертом Джентльменом [12]. R - це не лише мова, а й програмне забезпечення, найважливіші особливості наведені нижче:

- ефективна обробка даних та простий інструмент для зберігання результатів;
- набір операторів для обробки масивів, матриць та інших складних конструкцій;
- великі, послідовні, інтегровані інструменти збору для статистичного аналізу;
- численні графічні інструменти;
- проста і ефективна мова програмування, яка включає в себе безліч можливостей.

R - це мова з безліччю функцій, які виконують аналіз статистики та показують графічне відображення результатів, які візуалізуються у її вікні та дають можливість зберігати дані у різних форматах (jpg, png, bmp, eps або wmf у Windows, ps, bmp, pictex на Unix). R - мова, яка дозволяє користувачеві використовувати циклічні оператори, послідовно аналізувати кілька наборів даних. Також можна комбінувати різні статистичні функції для більш складного аналізу. Інтегроване середовище розробки, R Studio використовує мову R для розробки статистичних програм. У R можна написати програму та запустити код незалежно від будь-якої іншої комп'ютерної програми. R Studio, однак, для належного функціонування слід використовувати поряд з R. R може використовуватися без R Studio, але R Studio не може використовуватися без R [13-14].

- R - абсолютно безкоштовна програма. Програма R безкоштовна і, крім того,

ліцензується згідно з умовами ліцензії GNU General Public. Також велика кількість пакетів R може використовуватися за однією ліцензією в комерційних програмах.

- R популярний. IEEE публікує список найпопулярніших мов програмування щороку. Для мови, орієнтованої на домени, наприклад R, популярність важливіша, ніж для мови загального призначення, наприклад, C#. Це показує не тільки зростаючий інтерес до R як мови програмування, але й зростаючу популярність у таких сферах, як наука про дані та машинне навчання, де R зазвичай використовується.

- R працює на всіх платформах. Можна знайти пакети для установки R для всіх популярних платформ - Windows, Linux та Mac. Код R, який пишеться на одній платформі, без проблем можна перенести на іншу. Міжплатформна взаємодія є важливою особливістю сучасного комп'ютерного світу.

II Порівняльний аналіз R з альтернативними мовами програмування.

Як вже було сказано, R - не єдина мова, яку можна використовувати для статистичних обчислень та графіків. Нижче наведено порівняння трьох мов програмування за такими критеріями:

1. Доступність / вартість
2. Простота навчання
3. Можливості обробки даних
4. Графічні можливості
5. Інструментарій
6. Наявність обслуговування клієнтів

Порівняльний аналіз R, SAS, Python представлений у Табл.1.

Таблиця 1 – Порівняння з іншими мовами програмування

Мова	Вартість	Простота навчання	Можливості обробки даних	Графічні можливості	Інструментарій	Наявність обслуговування клієнтів
R	+	+/-	+	+	+	-
SAS	-	+/-	+	-	+	+
Python	+	+/-	+	+	+	-

1. SAS - комерційне програмне забезпечення. Поширене в приватних організаціях. Тоді як R і Python - абсолютно безкоштовні.

2. Всі порівнювані мови вимагають мінімального знання кодування. Звичайно існують ресурси для навчання, на веб-сайтах університетів є навчальні посібники і т.д.

3. Усі три мови мають можливості обробки даних та варіанти для проведення паралельних обчислень.

4. SAS має функціональні графічні можливості. Однак побудова графіків є складною і вимагає зрозуміти тонкощі використання пакету SAS Graph. R і Python мають розвинені графічні можливості. Існують численні пакети, які надають передові графічні можливості.

5. Усі три мови мають основні функції.

6. R та Python мають найбільші інтернет-спільноти, але не підтримують обслуговування клієнтів. Тож якщо виникають проблеми, то вирішувати їх треба самостійно. SAS підтримує функцію обслуговування клієнтів.

III Визначення можливостей застосування R для моделей та методів прогнозування цін на електричну енергію.

Найчастіше використовувані для побудови прогнозів на ринку електричної енергії моделі прогнозування: ARIMA (інтегрована модель авторегресії ковзкого середнього), SARIMA (сезонна інтегрована модель авторегресії ковзкого середнього), ARMAX (авторегресійна модель ковзкого середнього з екзогенними змінними), експоненційний метод

згладжування та його різновиди, баєсовські методи і т.д.

Серед основних можливостей R саме для прогнозування ціни на ринку електроенергії слід зазначити що це мова з відкритим кодом, багатофункціональна, сучасна за рахунок підтримки широкого штату розробників, які постійно вдосконалюють та оновлюють коди та пакети для обробки баз даних. R володіє безліччю статистичних функцій, які виконують аналіз даних за методиками всіх зазначених вище моделей, а також мають змогу графічно відображати результати [15-16], що безперечно є одним із найважливіших критеріїв для вибору саме цього інструменту при прогнозуванні цін на ринку електричної енергії. Також, безперечно, дуже важливим фактором для вибору цієї мови програмування є можливість роботи програми на всіх платформах, зберігання даних у різних форматах, відкриті енергетичні бази даних інших користувачів, це важливо для порівняльного аналізу.

Висновки

Після проведення порівняльного аналізу трьох мов програмування можна зробити висновок що R і Python, за обраними критеріями, мають однакові показники оцінки. Враховуючи відкритість коду обох мов, слід зазначити що саме R була розроблена для обробки статистичної інформації і саме R широко використовується для аналізу статистичних баз даних науковцями. Інтегроване середовище розробки R Studio має різні пакети для обробки баз даних, які корисні при прогнозуванні цін на ринку електроенергії: класифікація, регресія та оцінка розподілу, обчислення помилок моделей класифікації та регресії, оцінки втрат, оцінка та ефективність класифікаційних та регресійних моделей, обчислення точності, автокореляція та кореляція тощо.

Таким чином, представлені в даному матеріалі можливості програмного середовища R, незважаючи на їх стислість дозволяють показати, наскільки корисним буде для користувача їх застосування для прогнозування цін на електричну енергію на ринку електроенергії України.

References

1. Hyndman, R., Athanasopoulos, G., 2013. Forecasting: principles and practice. OTexts.
2. Forecasting: Principles and practice. 2016. Available from: Crossref
3. Kolassa, S., 2016. Evaluating predictive count data distributions in retail sales forecasting. International Journal of Forecasting 32 (3), 788–803.
4. Nikolaos Kourentzesa, Fotios Petropoulos, 2016. Forecasting with R. [Online]. Available: <https://kourentzes.com/forecasting/wp-content/uploads/2017/06/Forecasting-with-R-notes.pdf>
5. Kholmatova K. K., Kharkova O. A., Grjibovski A. M. Types of Research in Health Sciences. *Ekologiya cheloveka [Human Ecology]*. 2016, 1, pp. 57-64.
6. Buhovec A. G., Moskalev P. V., Bogatova V. P., Biryuchinskaya T. Y. *Statisticheskii analiz dannyh v sisteme R: uchebnoe posobie [Statistical data analysis using R: the textbook]*. Voronezh, 2010, 124 p.
7. Kabakov R. I. *R v deystvii. Analiz i vizualizatsiya dannykh v programme R [R in action: data analysis and visualization using R software]*, per. s angl. P. A. Volkov. Moscow, 2014, 588 p.
8. Mastickiy S. E., Shitikov V. K. *Statisticheskii analiz i vizualizatsiya dannykh s pomoshch'yu R [Data statistical analysis using R]*. Moscow, 2015, 496 p.
9. Crawley M. J. *The R Book*. 2nd ed. Wiley, 2013.
10. R: анализ и визуализация данных. Available from: <http://r-analytics.blogspot.ru> (accessed: 06.05.2020).
11. TIOBE Index for May 2020. Available from: <https://www.tiobe.com/tiobe-index/> (accessed: 06.05.2020).
12. Learn R Programming. The Definitive Guide, September [Online]. Available: <https://www.datamentor.io/r-programming/>; 2019.

13. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. (2017). URL <https://www.R-project.org/>
14. Official website, May [Online]. Available: <https://rstudio.com/>; 2020.
15. Nicholas J. Horton, Using R and R Studio for Data Management, Statistical Analysis, and Graphics. Second edition: book/ Nicholas J. Horton, [Online]. Available: <https://englianhu.files.wordpress.com/2016/01/using-r-and-rstudio-for-data-management-statistical-analysis-and-graphics-2nd-edit.pdf>
16. I. Setiawan, 2019. Time series air quality forecasting with R Language and R Studio. [Online]. Available: <https://iopscience.iop.org/article/10.1088/1742-6596/1450/1/012064/pdf>