



Прийнято 28.10.2025. Прорецензовано 23.12.2025. Опубліковано 29.12.2025.

УДК 658.5

DOI: 10.31471/1993-9981-2025-2(55)-101-107

ЗАСТОСУВАННЯ МЕТОДУ КЛАСТЕРИЗАЦІЇ ДЛЯ ПРОГНОЗУВАННЯ ЗНАЧЕНЬ ПРОДУКТИВНОСТІ ВІДЦЕНТРОВОГО НАГНІТАЧА

Гарасимів Т. Г.

Асистент кафедри комп'ютерних систем і мереж

Івано-Франківський національний технічний університет нафти і газу

76019, вул. Карпатська, 15, м. Івано-Франківськ, Україна

<https://orcid.org/0000-0003-1731-0286>

e-mail: taras.harasymiv@nung.edu.ua

Анотація. Проаналізовано підходи до прогнозування значень об'ємної продуктивності відцентрового нагнітача, що базуються на методах кластеризації. Основною перевагою запропонованих підходів над традиційними моделями прогнозування є їхня здатність враховувати стохастичний характер експериментальних даних та використовувати сучасні інформаційні технології, зокрема елементи нечіткої логіки, штучні нейронні мережі та генетичні алгоритми. Завдяки використанню багаторазових ітеративних процедур забезпечується поступове зниження похибки цільової функції до необхідного рівня, що сприяє підвищенню точності прогнозних результатів. Зокрема, розроблено метод прогнозування на основі алгоритму Fuzzy C-Means, який ґрунтується на концепції нечітких с-середніх. Його реалізацію здійснено в програмному середовищі Matlab. У межах цього підходу було встановлено 20 кластерів, а кількість ітераційних кроків складала до 100. Отримане середньоквадратичне відхилення становить 0.04, що свідчить про прийнятну точність моделі. Окрім цього, у роботі розглянуто альтернативний метод прогнозування на основі алгоритму субтрактивної нечіткої кластеризації. Побудова системи нечітких правил відбувалася автономно на основі експериментальних вимірювань із застосуванням моделі типу Сугено. Для кожного з досліджуваних об'єктів обчислювалося значення потенціалу, що характеризує ймовірність утворення кластера у його околі. Центри кластерів визначались ітераційно на основі точок із найбільшим потенціалом. У результаті середньоквадратичне відхилення, що було отримано для цього підходу, дорівнює 0.019, що підтверджує вищу точність прогнозування порівняно з методом Fuzzy C-Means. Апробація розглянутих методів здійснювалась на основі реальних експериментальних даних, що були зібрані під час експлуатації відцентрового нагнітача моделі 16 ГЦ2-396/53-78С у Долинському лінійному виробничому управлінні магістральних газопроводів у листопаді 2015 року.

Ключові слова: відцентровий нагнітач, об'ємна продуктивність, кластеризація, прогнозування, Fuzzy C-Means, субтрактивна нечітка кластеризація.

Запропоноване посилання: Гарасимів, Т. Г. (2025). Застосування методу кластеризації для прогнозування значень продуктивності відцентрового нагнітача. *Методи та прилади контролю якості*, 2(55), 101-107. doi: 10.31471/1993-9981-2025-2(55)-101-107

* Відповідальний автор



Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (<https://creativecommons.org/licenses/by/4.0/>)

Вступ

Сучасні підприємства та організації демонструють зростаючу потребу у впровадженні інформаційних систем, здатних не лише акумулювати та зберігати дані, а й забезпечувати ефективне управління, аналітичну обробку та прогнозування відповідно до концепцій і стандартів інтелектуального аналізу даних (Data Mining).

Традиційні методи прогнозування зазвичай ґрунтуються на побудові єдиної математичної моделі з обмеженим числом параметрів, яке визначається кількістю доступних експериментальних спостережень. Однак, на практиці формування такої узагальненої моделі є доволі складним завданням, а в окремих випадках – і зовсім недосяжним.

У зв'язку з цим актуальності набувають підходи до прогнозування, що враховують стохастичну природу даних та використовують сучасні методи кластеризації. Відомі нині підходи можна умовно класифікувати на три ключові групи згідно з теоретичними основами штучного інтелекту, які покладено в їх основу [1].

До першої групи належать нейронні мережі, що функціонують як універсальні адаптивні апроксиматори та здатні виявляти локальні закономірності у поведінці даних і точно їх моделювати. Друга група охоплює нечіткі логічні системи та гібридні нейронно-нечіткі моделі, які дозволяють формувати прогностичні структури з високою надійністю та інтерпретованістю логічних зв'язків.

Третя група представлена методами розподіленого інтелекту, до яких належать генетичні алгоритми, метод мурашиних колоній, а також підхід ройового інтелекту. Ці алгоритми, крім прогнозування, ефективно застосовуються для оптимізації параметрів, зокрема у завданнях визначення вагових коефіцієнтів у нейронних мережах.

Мета роботи – підвищення точності прогнозування значень об'ємної продуктивності відцентрового нагнітача шляхом розробки та дослідження методів кластеризації, зокрема алгоритмів нечітких с-середніх

(Fuzzy C-Means) та субтрактивної нечіткої кластеризації, із використанням сучасних інформаційних технологій (нечітка логіка, штучні нейронні мережі, генетичні алгоритми). Реалізація таких підходів дозволяє враховувати стохастичний характер експериментальних даних, забезпечувати зниження похибки цільової функції та досягати прийнятної точності прогнозних моделей.

Теоретичне обґрунтування

Станом на сьогодні методи кластеризації активно використовують для вирішення широкого кола задач, пов'язаних із моделюванням та прогнозуванням у різних галузях наукових досліджень. Застосування підходів, що базуються на часових рядах, є особливо актуальним у таких сферах, як економічне планування, торгівля, управлінські процеси, а також у проектуванні інтелектуальних систем підтримки прийняття рішень [2–4].

Основи теорії прогнозування з використанням інтелектуальних методів були закладені такими вченими, як Н. Вінер, Н. А. Колмогоров та Р. Калман [5]. У класичному підході до прогнозування для опису невизначеностей застосовуються засоби математичної статистики та теорії випадкових процесів.

Однак, з огляду на наявність хаотичної складової в більшості реальних процесів, виникає потреба у створенні таких методів прогнозування, які здатні об'єднувати набір моделей, адаптованих до різних закономірностей, що спостерігаються у експериментальних даних [6]. Саме тому значного поширення набули підходи, побудовані на основі кластеризаційних технологій, у яких формування моделей здійснюється відповідно до кластерів, утворених послідовностями експериментальних значень.

Для забезпечення стабільного та ефективного функціонування газоперекачувального агрегату (ГПА) в умовах реальної експлуатації, диспетчерський персонал компресорної станції (КС) здійснює моніторинг ряду теплотехнічних показників, зокрема: потужності та коефіцієнта корис-

ної дії (ККД) газотурбінного двигуна (ГТД), продуктивності агрегату, політропного ККД, вихідної потужності відцентрового нагнітача (ВЦН), приведеної потужності ГТД, коефіцієнтів технічного стану ГТД (за потужністю) та ВЦН (за політропним ККД), а також коефіцієнта режиму роботи ВЦН [7].

Для зниження ймовірності помилкових рішень у процесі керування технічними об'єктами застосовується прогнозування теплотехнічних параметрів ГПА, точність якого значною мірою залежить від обраних методів і доступних ресурсів. У випадках, коли досліджувані об'єкти функціонують за умов стохастичної невизначеності, класичні підходи до прогнозування ґрунтуються на використанні інструментарію математичної статистики та теорії випадкових процесів.

Для моделювання нелінійних залежностей у цій роботі застосовано методи нечіткої логіки, що базуються на формальних правилах, таких як «ЯКЩО – ТО». Якість побудованої нечіткої моделі визначається формою функцій належності та їх просторовим розташуванням. У разі використання нейро-нечітких систем процес налаштування параметрів моделі піддається автоматизації, при цьому адаптивна нейронна мережа виконує функцію оптимізації параметрів нечітких множин. У структурі моделі нечіткі множини слугують параметрами, а набір логічних правил і їх кількість визначають топологію самої моделі.

Для реалізації дослідження було використано експериментальні вимірювання, зафіксовані на компресорній станції «Долина» під час роботи відцентрового нагнітача природного газу марки 16 ГЦ2-396/53-78С у листопаді 2015 року.

Об'ємна продуктивність компресора, приведена до умов всмоктування, у одиницях м³/хв, розраховувалася згідно з такою формулою:

$$Q_{ec} = A_k \cdot \sqrt{\frac{\Delta P \cdot Z_1 \cdot R \cdot T_1}{P_1}}, \quad (1)$$

де ΔP – перепад тиску на конфузори, МПа; A_k – коефіцієнт конфузора, м²; R – газова стала газу, кДж/(кг·К); T_1 – температура на

вході нагнітача, К; Z_1 – коефіцієнт стисливості газу на вході нагнітача; P_1 – тиск газу на вході нагнітача, МПа.

На початковому етапі з експериментальних даних здійснюється формування нечітких правил шляхом застосування методу субтрактивної кластеризації. Далі, на другому етапі, здійснюється оптимізація параметрів нечіткої моделі за допомогою алгоритму ANFIS. Субтрактивна кластеризація виступає як швидкий і незалежний метод автоматичного генерування нечітких правил на основі наявних даних. Результати цього процесу можуть бути використані для побудови різних типів баз знань, зокрема сингтонних, Мамдані та Сугено моделей.

Основна ідея цього підходу полягає у попередньому визначенні точок у вхідному просторі, які можуть бути потенційними центрами кластерів. Для кожної такої точки обчислюється значення потенціалу, що відображає ймовірність утворення кластера в її околі. На наступному етапі ці точки аналізуються ітераційно, і серед них обираються ті, які мають максимальні значення потенціалу, формуючи центри кластерів.

Суттєвою перевагою використання кластеризаційних методів для синтезу нечітких моделей є їх здатність генерувати об'єктно-орієнтовані правила баз знань, які легко адаптуються до структури вихідних даних.

У програмному середовищі Matlab реалізовано два основних підходи до задачі нечіткої кластеризації:

- алгоритм FCM (Fuzzy C-Means) – вбудована функція, що реалізує метод кластеризації на основі нечітких с-середніх;
- субтрактивна нечітка кластеризація – метод, що дозволяє синтезувати нечіткі моделі на базі попереднього оцінювання щільності даних.

Алгоритм Fuzzy C-Means призначений для автоматизованого групування вхідних об'єктів у відповідні кластери, які трактуються як нечіткі множини. Межі між такими кластерами не є чітко визначеними, а ступінь належності кожного об'єкта визначається відстанню до центрів кластерів. Процес кластеризації відбувається ітера-

ційно: на кожному кроці уточнюється положення центрів кластерів та значення функцій належності відповідних елементів.

Припустимо, що маємо $K = 336$ експериментальних вхідних векторів x_k і N виявлених кластерів. Нехай ступінь належності будь-якого вектору x_k до відповідного кластеру c_j рівна $\mu_{jk} \in [0,1]$, де j – номер кластера, а k – номер вхідного вектора.

Запишемо умови нормування для μ_{jk} :

$$\begin{aligned} \sum_{j=1}^N \mu_{jk} &= 1, \forall k = 1, \dots, K; \\ 0 &< \sum_{k=1}^N \mu_{jk} \leq K, \forall j = 1, \dots, N. \end{aligned} \quad (2)$$

Метою роботи алгоритму є зведення до мінімуму суми всіх зважених відстаней між об'єктами та центрами кластерів.

$$\sum_{j=1}^N \sum_{k=1}^K (\mu_{jk})^q \|x_k - c_j\| \rightarrow \min, \quad (3)$$

де q – зафіксований параметр, який відомий перед ітераціями.

Надалі, враховуючи умову (3), розв'яжемо таку систему рівнянь:

$$\begin{cases} \frac{\partial}{\partial \mu_{jk}} \left(\sum_{j=1}^N \sum_{k=1}^K (\mu_{jk})^q \|x_k - c_j\| \right) = 0, \\ \frac{\partial}{\partial c_j} \left(\sum_{j=1}^N \sum_{k=1}^K (\mu_{jk})^q \|x_k - c_j\| \right) = 0. \end{cases} \quad (4)$$

Тоді розв'язання даної системи можна записати у наступному вигляді:

$$c_j = \frac{\sum_{k=1}^K (\mu_{jk})^q \cdot x_k}{\sum_{k=1}^K (\mu_{jk})^q}, \quad (5)$$

$$c_j = \frac{\sum_{k=1}^K (\mu_{jk})^q \cdot x_k}{\sum_{k=1}^K (\mu_{jk})^q}. \quad (6)$$

До застосування алгоритму FCM входить реалізація перелічених кроків:

Крок А. Ініціалізування. Необхідно вибрати наперед зазначену кількість кластерів N , визначити кількість ітерацій та зафіксований параметр q . Оскільки $2 < N < K$, то виходить, що $N=20$. Звідси бачимо, що шукана кількість ітерацій дорівнює 100, а фіксований параметр

дорівнює 1.5. Звідси бачимо, що початкова матриця належності дорівнює $U^{(0)} = (\mu_{jk})^{(0)}$.

Крок Б. Розрахунок величин центрів кластерів $c_j^{(t)}$. Да даному кроці здійснюємо розрахунок $c_j^{(t)}$, використовуючи формулу (5).

Крок В. Визначаємо значення функції належності μ_{jk} . Знаходимо $\mu_{jk}^{(t)}$, якщо $x_k \neq c_j$ та беремо до уваги вже відомі величини $c_j^{(t)}$. В іншому разі:

$$\mu_{jk}^{(t+1)} = \begin{cases} 1, & l = j, \\ 0. & \end{cases}$$

Крок Г. Завершення алгоритму або його заупинка. Зупинити даний алгоритм нечіткої кластеризації можна у разі виконання даної умови:

$$|U^{(t+1)} - U^t| \leq \varepsilon,$$

де $\varepsilon = 10^{-5}$ – точність обчислення алгоритму.

Результат застосування описаного алгоритму зображено на рисунку 1.

Для оцінки точності прогнозу значень об'ємної продуктивності нагнітача знайдено величину середньоквадратичного відхилення прогнозованих значень від знайдених значень об'ємної продуктивності за наведеною формулою [4]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1, n}^n (y_i - \hat{y}_i)^2}, \quad (7)$$

де $n=336$ – кількість експериментів; y_i – знайдені значення об'ємної продуктивності відцентрового нагнітача, м³/с; $\hat{y}_i$ – прогнозовані величини продуктивності із використанням алгоритму FCM, м³/с.

У результаті обчислень отримано значення середньоквадратичного відхилення, яке становить 0.04.

Для реалізації методу субтрактивної нечіткої кластеризації було використано інструментарій Fuzzy Logic Toolbox у середовищі Matlab. Застосування цього пакету дозволяє автоматично формувати нечітку базу знань на основі вхідних експериментальних даних із використанням моделі типу Сугено. У процесі виконання алгоритму генеруються правила, що відповідають ділянкам із максимальною густиною роз-

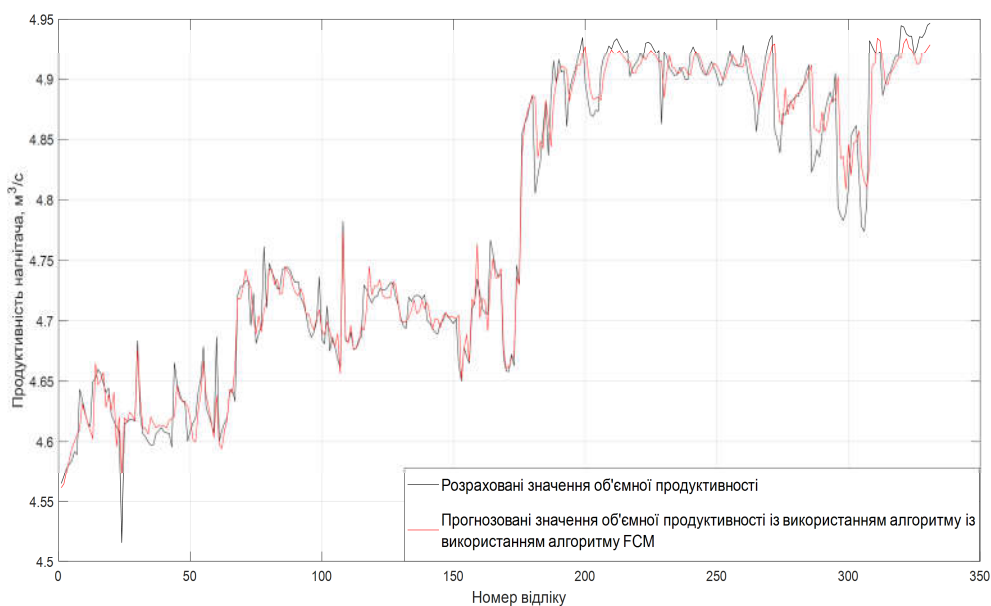


Рисунок 1 – Обчислені та прогнозовані величини об'ємної продуктивності відцентрового нагнітача із застосуванням алгоритму FCM

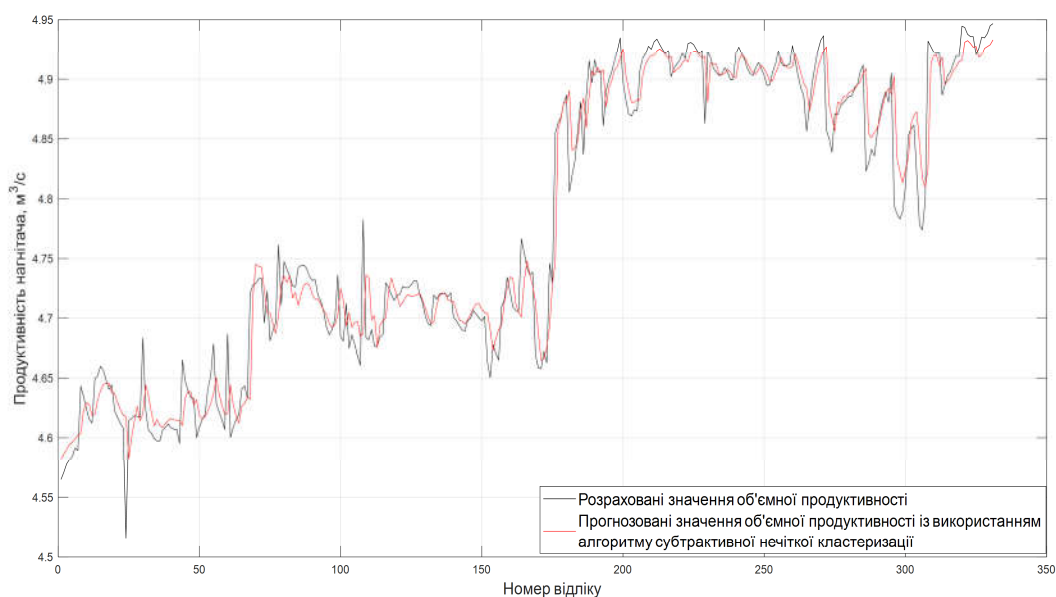


Рисунок 2 – Обчислені та прогнозовані величини об'ємної продуктивності відцентрового нагнітача із застосуванням алгоритму субтрактивної нечіткої кластеризації

міщення вихідних даних. Реалізація алгоритму здійснена за допомогою вбудованої функції *genfis2*.

Візуалізація результатів роботи алгоритму субтрактивної кластеризації наведена на рисунку 2.

Процедура витягування нечітких правил із експериментальних даних у межах функції *genfis2* умовно поділяється на два основні етапи. На першому з них за допо-

могою функції *subclust* визначається кількість кластерів, що відповідає кількості нечітких правил. На другому етапі, з використанням методу найменших квадратів (МНК), формується наслідкова частина кожного правила, тобто складова «...тоді...».

Одним із ключових параметрів функції є вектор *r*, який встановлює розмірність області нечіткого правила в напрямку ко-

жної координати. Зважаючи на те, що функція *subclust* виконує нормалізацію вхідних даних до одиничного гіперкуба, значення елементів вектора *г* повинні лежати в інтервалі $[0;1]$.

Варто зазначити, що вибір радіуса *г* істотно впливає на якість побудованої моделі. Якщо значення *г* є малим, то база нечітких правил буде більш деталізованою, проте така модель стає чутливою до шумів і статистичних викидів. Натомість, надто велике значення цього параметра може призвести до втрати частини потенційно корисних правил під час побудови моделі. Відповідно до рекомендацій, наведених у джерелі [8], у даному дослідженні обрано значення $г = 0.55$, а кількість ітерацій встановлено рівною 100.

Для перевірки точності прогнозованої моделі використано формулу (7), в результаті чого було отримано значення середньоквадратичного відхилення, що становить 0.019.

Висновки

У межах даного дослідження було запропоновано підхід до прогнозування значень об'ємної продуктивності нагнітача,

заснований на використанні алгоритму Fuzzy C-Means. За результатами моделювання визначено значення середньоквадратичного відхилення, яке становить 0.04.

Окрім того, реалізовано альтернативний метод прогнозування, що ґрунтується на субтрактивному алгоритмі нечіткої кластеризації. Формування нечітких правил здійснено автоматично на основі наявних експериментальних даних. Для побудови відповідної бази знань було використано модель типу Сугено.

Згідно з результатами розрахунків, значення середньоквадратичного відхилення, отримане для цього підходу, становить 0.019. Проведене порівняльне оцінювання підтвердило високу ефективність обох методів, зокрема — хорошу точність прогнозування об'ємної продуктивності нагнітача.

Подяки

Відсутні.

Конфлікт інтересів

Відсутній.

Список використаних джерел / References

1. Bezdek J.C., Keller J.M., Krishnapuram R., Pal N.R. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. Springer. 2005. 432 p. URL: <https://link.springer.com/book/10.1007/b106267>
2. Phu L., Anh D.T. Mostif-based method for initialization the K-means clustering for time series data. *Springer*. 2011. Vol. 7106. P. 11–20. doi: [10.1007/978-3-642-25832-9_2](https://doi.org/10.1007/978-3-642-25832-9_2)
3. Gonzalez H., Han J., Li X., Mythili V., Wang X. Adaptive mining of data streams using ensemble classifiers. *ACM SIGMOD Record*. 2012. № 41(1). P. 20–27.
4. Chatfield C. The Analysis of Time Series: An Introduction, Sixth Edition. CRC Press. 2009. 352 p.
5. Widiptura H., Kho H., Pears R., Kasabov N. A novel evolving clustering algorithm with polynomial regression for chaotic time-series prediction. *Neural Inf Process*. 2009. Vol. 5864. P. 114–121. doi: [10.1007/978-3-642-10684-2_13](https://doi.org/10.1007/978-3-642-10684-2_13)
6. Jang J.-S.R. ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems, Man, and Cybernetics*. 1993. Vol. 23(3). P. 665–685. doi: [10.1109/21.256541](https://doi.org/10.1109/21.256541)
7. Wang J., Chi D., Wu J., Lu H. Chaotic time series method combined with particle swarm optimization and trend adjustment for electricity demand forecasting. *Expert Systems with Applications*. 2011. Vol. 38. P. 8419–8429. doi: [10.1016/j.eswa.2011.01.037](https://doi.org/10.1016/j.eswa.2011.01.037)
8. Ahmed M.U., Begg R.K., Hassan M.M. A fuzzy approach for the evaluation of cluster validity indices. *Pattern Recognition*. 2012. Vol. 45(1). P. 336–346.

APPLICATION OF CLUSTERIZATION METHOD TO PREDICT CENTRIFUGAL SUPERFAN PRODUCTIVITY VALUES

Harasymiv T. H.

Assistant, Department of Computer Systems and Networks
Ivano-Frankivsk National Technical University of Oil and Gas
76019, 15 Karpatska St., Ivano-Frankivsk, Ukraine
<https://orcid.org/0000-0003-1731-0286>
e-mail: taras.harasymiv@nung.edu.ua

Abstract. This research analyzes approaches to predicting the volumetric efficiency of a centrifugal supercharger based on clustering methods. The main advantage of the proposed approaches over traditional forecasting models is their ability to take into account the stochastic nature of experimental data and use modern information technologies, in particular, elements of fuzzy logic, artificial neural networks and genetic algorithms. The use of multiple iterative procedures ensures a gradual decrease in the error of the objective function to the required level, which contributes to increasing the accuracy of forecast results. In particular, a forecasting method based on the Fuzzy C-Means algorithm was developed, which is based on the concept of fuzzy c-means. It was implemented in the Matlab software environment. Within this approach, 20 clusters were established, and the number of iteration steps was up to 100. The resulting standard deviation is 0.04, which indicates an acceptable accuracy of the model. In addition, the paper considers an alternative forecasting method based on the subtractive fuzzy clustering algorithm. The construction of a system of fuzzy rules was carried out autonomously on the basis of experimental measurements using a Sugeno-type model. For each of the studied objects, the potential value was calculated, which characterizes the probability of cluster formation in its vicinity. The cluster centers were determined iteratively based on the points with the highest potential. As a result, the standard deviation obtained for this approach is 0.019, which confirms the higher forecasting accuracy compared to the Fuzzy C-Means method. The testing of the considered methods was carried out on the basis of real experimental data collected during the operation of the centrifugal supercharger model 16 ГЦ2-396/53-78C in the Dolyna linear production department of main gas pipelines in November 2015.

Keywords: centrifugal supercharger, volumetric efficiency, clustering, forecasting, Fuzzy C-Means, subtractive fuzzy clustering.