

## APPLICATION OF THE NAIVE BAYES CLASSIFIER FOR CALCULATION AND PREDICTION OF OZONATION PROCESSES

V. Zakharov, O. Ustinov, Yu. Zmievskii, V. Myronchuk  
*National University of Food Technologies*

---

**Key words:**

*Ozone*  
*Organic pollution*  
*Naive Bayes classifier*  
*Ozone-gas composition*  
*Expert system*

---

**Article history:**

Received 13.09.2018  
Received in revised form  
26.09.2018  
Accepted 12.10.2018

---

**Corresponding author:**

V. Zakharov  
**E-mail:**  
npnuht@ukr.net

---

**ABSTRACT**

The paper presents an expert system developed (software) by the authors for calculating and prediction the ozonation process. The basis of the calculation algorithm was Naive Bayes classifier using the technologies of «machine learning». The program is written by using Python.

The aim of the work was to develop and define the necessary parameters, classes and basic principles for the formation of the training samples. Five classes were proposed that correspond to a certain percentage of dissolved ozone in the liquid phase, since this index is one of the main parameters in determining the effectiveness of ozonation. The forming of a training samples means creation of a set of events in which the set of values of the selected parameters correspond to a certain class. At this stage, the temperature of the treated liquid and the concentration of ozone in the ozone-gas composition were chosen for the parameters. The more events are presented in the training samples, the more precisely classification is.

The result of this work was the basal version of the expert system that allows to predict the process of dissolving ozone in a liquid for given values of temperature and concentration of ozone in the ozone-gas composition. The overall accuracy of the classification of the expert system was approximately 60%. Within the temperature range 0...15°C, the classification accuracy can be assumed to be 90%, with higher temperature values, it gradually decreases. Entering several new parameters can improve classification accuracy. The function of machine learning on the principle of «Supervised Learning» is also implemented in the program. It is implemented by an additional module, which, after the determination of dissolved ozone by user, requests to confirm the correctness of the results. If the user confirms the correctness of the classification, then the given event is entered into the training sample.

## ЗАСТОСУВАННЯ АЛГОРИТМУ НАЇВНОГО БАЄСОВОГО КЛАСИФІКАТОРА ДЛЯ РОЗРАХУНКУ ТА ПРОГНОЗУВАННЯ ПРОЦЕСІВ ОЗОНУВАННЯ

В.В. Захаров, О.А. Устінов, Ю.Г. Змієвський, В.Г. Мирончук

Національний університет харчових технологій

У статті представлено розроблену авторами експертну систему (програмне забезпечення) для розрахунку та прогнозування процесу озонування. За основу алгоритму розрахунку було взято наївний Баєсів класифікатор із застосуванням технологій «машинного навчання». Програму написано у середовищі для програмування Python.

Розроблено і визначено необхідні параметри, класи та основоположні принципи формування навчальної вибірки. Запропоновано п'ять класів, які відповідають певному відсотку розчиненого озону в рідкій фазі, оскільки цей показник один із головних параметрів при визначенні ефективності озонування. Принцип формування навчальної вибірки полягає у створенні набору подій, у яких сукупність значень обраних параметрів відповідає певному класу. На цьому етапі за параметри було обрано температуру оброблюваної рідини і концентрацію озону в озono-газовій суміші. Чим більше подій відбувається у навчальній вибірці, тим точніше проходить класифікація.

Результатом дослідження стала базова версія експертної системи, яка дає змогу спрогнозувати процес розчинення озону в рідині при заданих значеннях температури та концентрації озону в озono-газовій суміші. Загальна точність класифікації експертної системи склала приблизно 60%. У межах значень температури 0...15°C точність класифікації можна вважати 90%, при вищих значеннях температури вона поступово зменшується. Підвищити точність класифікації можна шляхом введення кількох нових параметрів. Також у програмі реалізовано функцію машинного навчання за принципом «Supervised Learning», завдяки додатковому модулю, який після визначення користувачем розчиненого озону видає запит на підтвердження правильності отриманих результатів. Якщо користувач підтверджує правильність класифікації, то задана подія вноситься до навчальної вибірки.

**Ключові слова:** озон, органічне забруднення, наївний Баєсів класифікатор, озono-газова суміш, експертна система.

**Постановка проблеми.** Застосування процесу озонування для окислення органічних домішок та їх подальшого видалення на сорбційних фільтрах має значну перспективу використання у харчовій промисловості. Використовуючи цей процес, можна видаляти небажані органічні домішки з оброблюваних розчинів і забезпечувати мікробіологічну чистоту технологічного обладнання [1—4]. Значною перевагою озонування є його екологічність і безпечність для харчових виробництв [1; 4—5]. За його допомогою у харчовій промисловості вже обробляють сировину й тару в овочево-фруктової,

рибній, м'ясній і зерновій галузях для знищення мікроорганізмів, які призводять до появи плісняви на поверхні сировини під час її зберігання. Озоном обробляють молоко, соки, сидр, пиво, міцні алкогольні напої з метою підвищення строків їх придатності [1; 3—7]. На сьогодні існує вкрай обмежена кількість методів, які дають змогу з великою точністю прогнозувати ефективність зазначених процесів, тому розробка нових є актуальним завданням.

**Аналіз останніх публікацій і досліджень.** Авторами попередньо проведені експерименти з озонування нанофільтраційного пермеату молочної сироватки — одного з перспективних напрямів застосування озонування. Було встановлено раціональні параметри процесу [8], проте впродовж дослідження виявилось, що ефективність процесу озонування залежить від багатьох факторів. Прогнозування та суворе дотримання рекомендованих параметрів можуть зробити процес озонування економічно вигідною технологією. Більшість дослідників, які вивчають озонування, відзначають необхідність проведення допроектних досліджень, проте це не завжди можна виконати без суттєвих затрат на лабораторне, експериментальне та напіввиробниче обладнання. А в деяких випадках необхідно лише перевірити саму можливість застосування процесу озонування. Зазвичай, у харчовій промисловості для таких цілей існують різноманітні розроблені моделі, спеціальні програми та алгоритми розрахунків. Проте сумарна складність усіх процесів, які відбуваються під час озонування, не дає змоги виробити єдиний та, головне, міжнародно апробований та визнаний спосіб моделювання процесу озонування. До того ж в існуючих моделях озонування не обходиться без коефіцієнтів або параметрів, які заходяться емпіричним шляхом, тобто проведенням прямого експерименту.

Висновком багатьох професіоналів у сфері озонування є те, що рекомендації щодо застосування озонування необхідно давати з урахуванням власного досвіду або після проведення попередніх експериментальних досліджень.

Проте розвиток техніки, особливо інформаційних технологій, дає змогу у XXI столітті по-новому застосовувати відомі математичні та статистичні прийоми. Тож у статті пропонується розроблений авторами метод визначення параметрів процесу озонування та його прогнозування, який заснований на найвньому Баєсовому класифікаторі із застосуванням технологій «машинного навчання» (англ. Machine Learning).

**Мета статті:** розроблення доступного для широкого спектра спеціалістів та зрозумілого способу розрахунку процесу озонування та його реалізація у вигляді програмного забезпечення.

**Методи та обладнання.** Програма для розрахунку процесу озонування була написана на мові програмування Python (Пайтон) версії 2.7.15. Python — потужна мова програмування, має ефективні структури даних високого рівня, простий і водночас ефективний підхід до об'єктно-орієнтовного програмування. Ця мова програмування має зручний та інтуїтивно зрозумілий синтаксис, динамічну обробку типів, вбудований інтерпретатор і велику кількість різноманітних модулів для широкого спектра задач, що робить її придатною для розробки прикладних програм [9].

В основу розрахунку був покладений наївний Баєсів класифікатор [10] — це один з імовірнісних класифікаторів, який визначає належність елемента вибірки до одного з класів. В основі роботи цього класифікатора лежить теорема Баєса та припущення (наївне) незалежності усіх змінних [10].

Припущення «наївності» Баєсового класифікатора полягає в тому, що, незважаючи на незалежність або залежність параметрів один від одного та їх зв'язків між собою, вважається, що при визначенні класу кожний з цих параметрів робить свій окреий і незалежний ні від чого внесок.

Переваги наївного Баєсівського класифікатора полягають у відносній простоті та наочності методу, що дає змогу досить швидко реалізовувати його в програмному кодї. При цьому на точність класифікація суттєво впливає правильний вибір параметрів і класів, які формують навчальну вибірку. Алгоритм працює швидко та не потребує великих затрат ресурсів комп'ютера.

**Результати та їх обговорення.** Серед багатьох параметрів і показників процесу озонування одну з головних ролей відіграє кількість розчиненого озону в оброблюваному розчині  $RO$ . Це пов'язано з тим, що в реакцію з обраними для обробки речовинами у розчинї вступає саме розчинений у рідкій фазї озон. Головним чином ця величина залежить від  $pH$  розчину, його температури ( $T$ ), вмісту речовин, здатних окислитися озоном ( $M$ ), та концентрації озону в озono-газовій суміші ( $C_0$ ):

$$RO = f(pH, T, M, C_0). \quad (1)$$

Оскільки на цьому етапі необхідно було сформува ти вибірку даних і принципи роботи з нею, було прийнято рішення використати температуру та концентрацію озону в озono-газовій суміші як вихідні (початкові) параметри.

Вищезазначені параметри були взяті з багаторічного досвіду інженерів та науковців французької компанії Degremont, яка майже 80 років спеціалізується на виробництві питної води, водопідготовці та водовідведенні [11; 12] На рис. 1 представлені вихідні данні у вигляді графіків залежності концентрації озону в рідкій фазї залежно від його концентрації у газoвій фазї для різних значень температур.

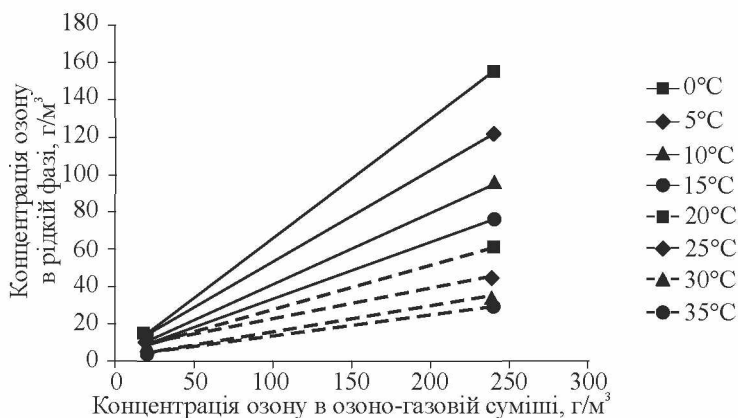


Рис. Набір даних для написання алгоритму

*Таблиця 1. Поділ діапазону значень розчинності озону на класи*

Клас	А	В	С	D	Е
Діапазон значень, %	$80 \leq Y \leq 100$	$60 \leq Y < 80$	$40 \leq Y < 60$	$20 \leq Y < 40$	$0 \leq Y < 20$

Параметри відповідають концентрації озону, що розчинилась у рідкій фазі  $Y$ . Було обрано п'ять класів: «А», «В», «С», «D», «Е». Їх сформовано відповідно до різних діапазонів значень  $Y$  (табл. 1).

При розподілі класів за основу взято клас С (40...60% розчиненого озону) — така область значень розчиненого озону в рідкій фазі вважається доцільною для використання з точки зору балансу між затратами на озонування та отриманим ефектом. Для інших класів було обрано крок у 20%, де В клас — вище за середній (доцільний); А — ідеальний, на цьому етапі цей клас не застосовується у навчальній вибірці, оскільки може бути досягнутий тільки при використанні додаткових засобів для розчинення озону (спеціальні конструкційні рішення для ректорів або досягнення певних технологічних показників при озонуванні); D — доцільний лише у разі відсутності інших альтернатив, окрім застосування озонування; Е — недоцільний, оскільки економічні затрати на вироблення озону значно перевищуватимуть отриманий ефект.

Принцип формування класів і параметрів навчальної вибірки представлено в табл. 2.

*Таблиця 2. Формування класів і параметрів навчальної вибірки*

№	T, °C	X, г/м <sup>3</sup>	Y, %	Клас
1.	0	240	74	В
2.	5	120	47	С
3.	0	70	66	В
4.	10	200	40	С
5.	35	20	18	Е
6.	20	118	13	Е
7.	0	50	67	В
8.	15	29	38	D
...	...	...	...	...

На основі зазначеної навчальної вибірки програма прогнозує найбільш імовірний діапазон (в нашій термінології «клас»), що відповідає проценту розчиненого озону в реакторі, використовуючи застосований алгоритм.

Принцип дії алгоритму наївного Баєсового класифікатора такий:

1. Користувач вводить запит, вказуючи температуру  $T_0$  [°C] та концентрацію озону в газовій фазі  $x_0$  [г/м<sup>3</sup>].

2. Розраховується апіорна ймовірність для кожного класу:

$$P(C_j) = \frac{N(C_j)}{N}, \quad (2)$$

де  $N(C_j)$  — кількість записів у навчальній вибірці, що відповідають цьому класу  $C_j$ ;  $N$  — об'єм навчальної вибірки,  $C_j = \{C_1, C_2, C_3, C_4, C_5\}$  — сукупність класів.

3. Перевірка апіорних імовірностей за класами:

$$\sum P(C_i) = 1. \quad (3)$$

4. Визначаємо апіорну ймовірність предикторів:

$$P(x = x_0) = \frac{N(x_0)}{N}; \quad (4)$$

$$P(T = T_0) = \frac{N(T_0)}{N}, \quad (5)$$

де  $N(x_0)$ ,  $N(T_0)$  — частоти значень параметрів  $x = x_0$  та  $T = T_0$  відповідно.

Для уникнення проблеми «нульових частот» було введено поправки  $\Delta x = \pm 0,5$  та  $\Delta T = \pm 2,5$  в межах яких враховується збіг із записом у навчальній вибірці.

5. Розрахунок «правдоподібності» для кожного параметра у всіх класах:

$$P(x = x_0 | C_j) = \frac{N(x_0 | C_j)}{N_j}; \quad (6)$$

$$P(T = T_0 | C_j) = \frac{N(T_0 | C_j)}{N_j}, \quad (7)$$

де  $N(x_0 | C_j)$ ,  $N(T_0 | C_j)$  — частоти значень параметрів при певному класі для  $x_0$  і  $T_0$  відповідно;  $N_j$  — об'єм вибірки для певного класу.

6. За формулою Байєса для кожного класу розраховуємо ймовірність того, що набір параметрів  $x = x_0$  та  $T = T_0$  відповідає певному класу  $C_j$ :

$$P(C_j | x = x_0, T = T_0) = \frac{P(x = x_0 | C_j) \cdot P(T = T_0 | C_j) \cdot P(C_j)}{P(x = x_0) \cdot P(T = T_0)}. \quad (8)$$

7. Класифікатор визначає найбільшу ймовірність належності до класу і видає прогноз користувачу.

8. Якщо користувач підтверджує факт того, що класифікація пройшла вірно, то програма доповнює навчальну вибірку записом (значення  $x_0$ ,  $T_0$  відповідають певному класу  $C_j$ ).

Восьмий пункт — це запропонована авторами реалізація машинного навчання програми. Її принцип «Supervised Learning» — навчання «з учителем», тобто задаються приклади з «вірними відповідями», які додаються до навчальної вибірки. За рахунок реалізації цієї функції і відбувається подальше навчання програми (розширення її початкової вибірки). Це дасть змогу розширювати та покращувати якість класифікації, що в результаті і зробить із програми потужний інструмент для експертної системи.

У табл. 3 наведено результати роботи розробленої експертної системи. Реальні розраховані дані кількості розчиненого озону  $RO$  було порівняно з прогнозом класифікатора.

**Таблиця 3. Результати порівняння вимірної розчинності озону з прогнозом класифікатора**

№ п/п	$T, ^\circ\text{C}$	$C_0, \text{г/м}^3$	Вимірний $RO, \%$	Клас за класифікатором	Достовірність
1	0	91	65,68	B	+
2	2	200	59,92	C	+
3	10	141	40,17	C	+
4	12	100	38,00	C	-
5	0	24	18,07	B	+
6	0	40	70,11	B	+
7	5	233	51,00	C	+
8	25	50	30,00	C	-
9	15	175	31,14	C	-
10	20	120	27,86	C	-
Сумарна достовірність класифікатора					60%

На цьому етапі розробки достовірність класифікатора складає 60%. З вибірки даних видно, що при температурах від 0 до 15°C достовірність класифікації складає майже 90% і поступово знижується при підвищенні температури. Зміна концентрації озону в озono-газовій суміші впливає на точність класифікатора меншою мірою. Точність класифікації можна підвищити у перспективі, додавши ще кілька параметрів.

### **Висновки**

У результаті проведеного дослідження було розроблено базову версію експертної системи для створення в подальшому потужного комплексу розрахунків і підбору необхідних параметрів роботи станцій озонування. В перспективі отриманий комплекс з допомогою машинного навчання може стати потужним інструментом у руках інженерних спеціалістів і науковців у сферах застосування технологій озонування. Перевагами комплексу є простота у використанні, можливість його самовдосконалення шляхом застосування машинного навчання, що поступово буде збільшувати точність та ефективність запропонованої експертної системи (програмного забезпечення).

### **Література**

1. Pandiselvam R., Sunoj S. Manikantan M.R., Kothakota A., Hebbar K.B. Application and Kinetics of Ozone in Food Preservation. *Ozone: Science & Engineering*. 2017. No. 39(2). P. 115—126.
2. Cullen P.J. Tiwari B.K., O'Donnell C.P., Muthukumarappan K.K. Modelling approaches to ozone processing of liquid foods. *Trends in Food Science & Technology*. 2009. No. 20. P. 125—136.
3. Pereira A.M., Faroni L.R.D., Silva A.G. da S.Jr., Sousa A.H., Paes J.L. Economical viability of ozone use as fumigant of stored corn grains. *Engenharia na Agricultura*. 2008. No. 16(2). P. 144—154.
4. Sequential Treatment of Tequila Industry Vinasses by Biopolymer-based Coagulation/Flocculation and Catalytic Ozonation / H. Ferral-Pérez, H. Torres Bustillos, L. Méndez та ін.; *Ozone: Science & Engineering*. 2016. No. 38. P. 279—290.

5. Karaca H. Use of Ozone in the Citrus Industry. *Ozone: Science & Engineering*. 2010. No. 32. P. 122—129.
6. Shigezo N., Takahara H. Ozone Contribution in Food Industry in Japan. *Ozone: Science and Engineering*. 2006. No. 28. P. 425—429.
7. Desvignes C. Changes in common wheat grain milling behaviour and tissue mechanical properties following ozone treatment / C. Desvignes, M. Chaurand, M. Dubois, A. Sadoudi, J. Abecassis, V. Lullien-Pellerin; *Journal of Cereal Science*. 2008. No. 47. P. 245—251.
8. Захаров В.В., Змієвський Ю.Г., Білецька І.М., Мирончук В.Г. Озонування рідин молочної промисловості. *Наукові праці Національного університету харчових технологій*. Київ. 2017. Том 23, № 5, Ч.1. С. 124—130.
9. Мова програмування Python, корпорації Python Software Foundation [US] URL: <https://www.python.org/downloads/release/python-2715/> www.python.org/
10. Langseth H., Nielsen T.D. Classification using Hierarchical Naive Bayes models. *Mach Learn*. 2006. No. 63. P. 135—159.
11. Degremont. Технический справочник по обработке воды. Том 1. / L. Andriamirado, D. Asensi, S. Baig та ін.; Санкт-Петербург: Новый журнал, 2007. 878 с.
12. Degremont. Технический справочник по обработке воды. Том 2. / L. Andriamirado, D. Asensi, S. Baig та ін.; Санкт-Петербург: Новый журнал, 2007. 920 с.