

DOI: <https://doi.org/10.32689/2523-4536/59-8>
УДК 334.758

Квашук Д. М.

кандидат економічних наук,
Національний авіаційний університет
ORCID: <https://orcid.org/0000-0002-4591-8881>

Kvashuk Dmytro

Candidate of Economic Sciences,
National Aviation University

ЗАСТОСУВАННЯ МЕТОДУ К-НАЙБЛИЖЧИХ СУСІДІВ В ПРОЦЕСІ ІНФОРМАЦІЙНО-АНАЛІТИЧНОГО ЗАБЕЗПЕЧЕННЯ ЕКОНОМІЧНОЇ БЕЗПЕКИ МОЛОКОПЕРЕРОБНИХ ПІДПРИЄМСТВ

APPLICATION OF THE METHOD OF K-NEAREST NEIGHBORS IN THE PROCESS OF INFORMATION AND ANALYTICAL SUPPORT OF ECONOMIC SECURITY OF DAIRY ENTERPRISES

У статті досліджено можливості класифікації загроз для молокопереробних підприємств з використанням методу *k*-найближчих сусідів. Метою застосування даного методу є вирішення задач з оцінювання постачальників сировини за окремими категоріями надійності. Особливості економічної діяльності молокопереробних підприємств обумовлені постійним пошуком якісної сировини. Звичайно, що перед її використанням на підприємствах застосовують ряд перевірок, але такі перевірки досить витратні і не завжди проводяться в силу різних причин. Досліджений в роботі метод класифікації інформації, може бути використаний із застосуванням технологій машинного навчання. Для його апробації було використано програмний код, який включає в себе такі бібліотеки, як *math*, *random*, *pylab*, *numpy*, *matplotlib*. Для класифікації постачальників сировини, було створено тестову вибірку даних, після чого обчислено відстань до кожного з об'єктів навчальної вибірки. Відібрано *k* об'єктів навчальної вибірки, відстань до яких є мінімально. А також визначено класи постачальників сировини. Частина уваги приділено особливостям постачальників сировини для молокопереробних підприємств. На основі дослідження особливостей діяльності постачальників сировини для молокопереробних підприємств встановлено кількісні критерії їх надійності. Так, відповідні критерії були оцінені із застосуванням методу *k*-найближчих сусідів, що дозволило сформувати класифікацію постачальників сировини. Разом з тим, у статті висвітлено характеристики інформаційно-аналітичного забезпечення економічної безпеки молокопереробних підприємств, виділено ряд ключових особливостей такої діяльності.

Ключові слова: метод *k*-найближчих сусідів, інформаційно-аналітичне забезпечення, економічна безпека, молокопереробні підприємства.

The article investigates the possibilities of classification of threats to dairy enterprises using the method of *k*-nearest neighbors. The purpose of this method is to solve problems of evaluating suppliers of raw materials for certain categories of reliability. Features of economic activity of milk processing enterprises are caused by constant search of high-quality raw materials. Of course, before using it, enterprises use a number of inspections, but such inspections are quite expensive and are not always carried out for various reasons. The method of information classification studied in the work can be used with the use of machine learning technologies. For its approbation the program code which includes such libraries as *math*, *random*, *pylab*, *numpy*, *matplotlib* was used. To classify suppliers of raw materials, a test sample of data was created, after which the distance to each of the objects of the training sample was calculated. Selected *k* objects of the training sample, the distance to which is minimal. And also the class of the supplier of raw materials which meets criteria of reliability is defined. Part of the attention is paid to the characteristics of suppliers of raw materials for dairy enterprises: conditions for keeping cows, approaches to the collection of raw materials, storage of raw materials and their transportation. Based on the study of the peculiarities of the suppliers of raw materials for dairy enterprises, quantitative criteria for their reliability have been established. Thus, the relevant criteria were evaluated using the method of *k*-nearest neighbors, which allowed to form a classification of suppliers of raw materials. At the same time, the article highlights the characteristics of information and analytical support of economic security of dairy enterprises, highlights a number of key features of such activities.

Keywords: *k*-nearest neighbors method, information-analytical support, economic security, milk processing enterprises.

Постановка проблеми. Безпека молокопереробних підприємств залежить від багатьох чинників як зовнішніх, так і вну-

трішніх. Одним із головних можна вважати якість сировини, тобто молочних продуктів, яка залежить від постачальників, що несуть

відповідальність за стан отриманого ними молока.

Постачання неякісних продуктів, несе загрозу для молокопереробних підприємств. Тому підприємці намагаються попередити таку загрозу шляхом прийняття потрібних для цього управлінських рішень. Такі рішення, необхідно приймати на основі реальних інформаційних показників, що певним чином відображають стан захищеності підприємства в умовах загроз, ризиків та небезпек, оскільки вирішення нагальних потреб підприємства в сучасних умовах потребує значної інформаційної підтримки. Механізми протидії загрозам та ризикам мають бути превентивного характеру, тому що загрозу легше попередити, ніж потім ліквідувати її наслідки.

Водночас забезпечення належним чином інформаційної підтримки стратегічних рішень у сфері економічної безпеки молокопереробних підприємств, істотно знижує впливи негативних факторів на їх діяльність, як з боку зовнішнього, так і внутрішнього середовища. Для цього необхідно визначити критерії небезпеки, оцінити її кількісно та класифікувати.

Метою статі є дослідження можливостей застосування методу k -найближчих сусідів для класифікації постачальників сировини молокопереробним підприємствам на предмет надійності.

Виклад основного матеріалу дослідження. Збільшення ризиків для молокопереробних підприємств обумовлене багатьма факторами, одним із яких є недобросовісність постачальників сировини. В таких умовах існує необхідність їх класифікувати, з метою виявлення окремих категорій постачальників, які мають високий ступінь надійності.

Потреба у використанні простого та надійного методу класифікації постачальників сировини існує вже давно, проте способи її вирішення мають різноплановий характер. Існують підходи до класифікації постачальників за допомогою матриці Бостонської консультативної групи [1], інші вчені отримують результат обробляючи великі дані [2]. В значній мірі застосовуються різні моделі прийняття рішень [3].

Запропонований у статті метод для вирішення задач із класифікації також набирає широкої популярності, особливо в областях машинного навчання.

Принцип дії методу k -найближчих сусідів відноситься до метричних методів класифікації. Так, для встановлення належності постачальника сировини d до окремої категорії, програмний класифікатор порівнює d з усіма аналогічними постачальниками за рядом кри-

теріїв з навчальної вибірки L , тобто для кожного $d_z \in L$ обчислюється відстань $p(d_z, d)$, після чого із навчальної вибірки вибираються k постачальників, які є найближчими до d . Так, згідно з методом k найближчих сусідів, постачальник d може вважатися належним того класу, який є найбільш поширеним серед сусідів даного постачальника, тобто для кожного класу обчислюється функція ранжирування (рис. 1):

$$CV(d) = \sum_{d_z \in L_k(d)} p(d_z, d) F(d_z, c_i), \quad (1)$$

де: $L_k(d)$ – найближчі k постачальники із вибірки L до постачальника d ;

$F(d_z, c_i)$ – відомі величини, постачальники які вже класифіковані за категоріями з навчальної вибірки.

Таким чином даний метод може бути застосований для класифікації надійності постачальників сировини для молокопереробних підприємств в процесі інформаційно-аналітичного забезпечення економічної безпеки.

Його недоліками можна вважати те, що існує значна залежність результатів класифікації від вибраної вибірки навчальних даних, значний час для проведення машинного аналізу, оскільки повного перебору вибірки навчальної даних.

До переваг слід віднести просту програмну реалізацію, стійкість алгоритму до аномальних викидів вихідних даних, можливість оновлення навчальної вибірки без перенавчання класифікатора даних [4].

Припустимо, що існують три категорії постачальників, оцінка яких може здійснюватися за такими критеріями: час існування на ринку, кількість скарг на постачальника.

Практична реалізація даного методу може бути сформована за наступним алгоритмом:

- для визначення категорії надійності постачальника необхідно обчислити відстань до кожного з об'єктів навчальної вибірки;
- відібрати k об'єктів навчальної вибірки, відстань до яких є мінімальною;
- визначити клас постачальника в залежності від його найближчих за критеріями сусідів (постачальників).

Для реалізації таких задач можна використати запропонований програмний код представлений спільнотою ІТ спеціалістів Хабр [5].

Так, використовуючи програмні бібліотеки `math`, `random`, `pylab`, `numpy`, `matplotlib`, для проведення експерименту, можна створивши умовний двовимірний масив даних, в якому випадковим чином на ділянці від 0 до 5 по кожній з осей вибирається тривалість діяльності постачальників, та кількість скарг.

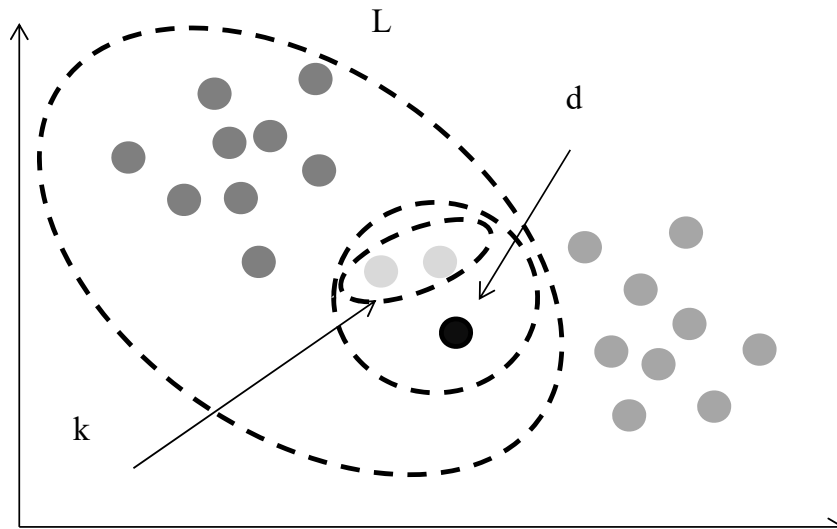


Рис. 1. Принцип реалізації методу k -найближчих сусідів

Для виділення проведення умовної класифікації в процесі формування вибірки було застосовано середньоквадратичне відхилення 0.5, а також за правилом трьох сигм [6, с. 31], було створено три класи даних.

Після ініціалізації зазначених вище програмних бібліотек із застосуванням мови програмування Python, генерація такої вибірки може бути створена з використанням наступного коду:

```
from matplotlib.colors import ListedColormap
import pylab as plt
import random
import math
import numpy as numpy
def GenDat (k, ClassData):
    data = []
    for classNum in range(ClassData):
        #Choose random center of 2-dimensional gaussian
        Cx,Cy=random.random()*5.0,random.random()*5.0
        #Choose k random nodes with RMS=0.5
        for rNum in range(k):
            data.append([[random.gauss(Cx,0.5),random.gauss(Cy,0.5)], classNum])
    return data
def shDat (nClass, itemClass):
    learnDat=GenDat (itemClass, nClass)
    classColormap=ListedColormap(['#FF0000', '#00FF00', '#0000CD'])
    plt.scatter([learnDat[i][0][0] for i in range(len(learnDat))],
                [learnDat[i][0][1] for i in range(len(learnDat))],
                c=[learnDat[i][1] for i in range(len(learnDat))],
                cmap=classColormap)
    plt.show()
shDat (3, 80)
```

Результат візуалізації представлено на (рис. 2), де видно розмежування між категоріями умовних постачальників сировини.

Маючи вибірку даних їх слід розділити на навчальну вибірку та тестову:

```
def TraintTest (data, TestPer):
    trDat = []
    tesDat = []
    for row in data:
        if random.random() < TestPer:
            tesDat.append(row)
        else:
            trDat.append(row)
    return trDat, tesDat
```

Маючи навчальну вибірку, алгоритм класифікації можна представити наступним чином:

```
def TraintTest (data, TestPer):
    trDat = []
    tesDat = []
    for row in data:
        if random.random() < TestPer:
            tesDat.append(row)
        else:
            trDat.append(row)
    return trDat, tesDat
def Knnclass (trDat, tesDat, k, numClass):
    #Euclidean distance between 2-dimensional point
    def distans (a, b):
        return math.sqrt((a[0] - b[0])**2 + (a[1] - b[1])**2)
    tesLab = []
    for tesP in tesDat:
        #Claculate distances between test point and all of
        the train points
        testDist = [ [distans(tesP, trDat[i][0]), trDat[i][1]] for
        i in range(len(trDat))]
        #How many points of each class among nearest K
        stat = [0 for i in range(numClass)]
        for d in sorted(testDist)[0:k]:
            stat[d[1]] += 1
        #Assign a class with the most number of occurences
        among K nearest neighbours
        tesLab.append( sorted(zip(stat, range(numClass)),
        reverse=True)[0][1] )
    return tesLab
```

Для визначення відстані між об'єктами було застосовано евклідову відстань.

Візуалізувати класифікаційні ознаки за окремими категоріями постачальників можна з використанням наступного коду (рис. 3):

```
def showDatMesh (nClass, nClassItem, k):
    #Generate a mesh of nodes that covers all train cases
```

```
def genTestMesh (trDat):
    x_min = min( [trDat[i][0][0] for i in
    range(len(trDat))] ) - 1.0
    x_max = max( [trDat[i][0][0] for i in
    range(len(trDat))] ) + 1.0
    y_min = min( [trDat[i][0][1] for i in
    range(len(trDat))] ) - 1.0
```

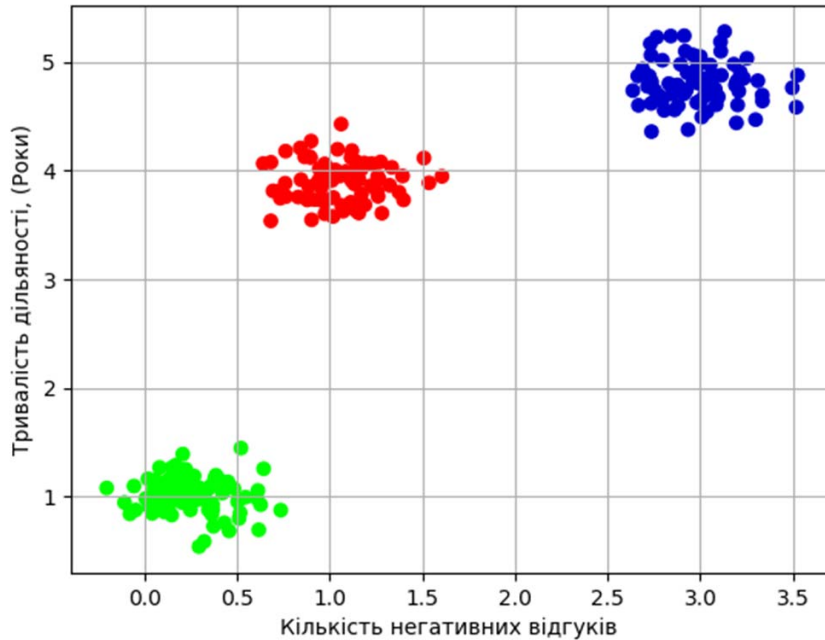


Рис. 2. Візуалізація створеної вибірки

Джерело: озроблено із застосування програмно-графічної бібліотеки matplotlib [7]

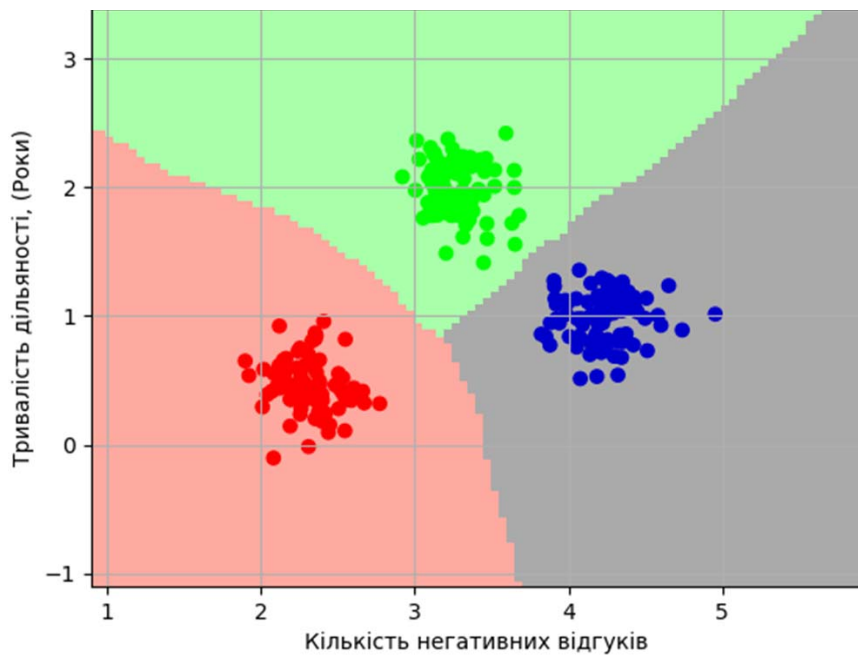


Рис. 3. Визначення категорій постачальників сировини для молокопереробних підприємств з використанням методу k-найближчих сусідів

Джерело: розроблено із застосування програмно-графічної бібліотеки matplotlib [7]


```

    y_max = max( [trDat[i][0][1] for i in range
(len(trDat))] ) + 1.0
    h = 0.05
    tX, tY = np.meshgrid(np.arange(x_min, x_max, h),
                        np.arange(y_min, y_max, h))
    return [tX, tY]
trDat = GenDat (nClassItem, nClass)
Tmesh = genTestMesh (trDat)
testMeshLabels = Knnclass (trDat, zip(Tmesh[0],
ravel(), Tmesh[1].ravel()), k, nClass)
classColormap = ListedColormap(['#FF0000',
'#00FF00', '#0000CD'])
testColormap = ListedColormap (['#FFAAAA',
'#AAFFAA', '#AAAAAA'])
plt.pcolormesh(Tmesh[0],
Tmesh[1],
np.asarray(testMeshLabels).
reshape(Tmesh[0].shape),
cmap=testColormap)
plt.scatter([trDat[i][0][0] for i in range(len(trDat))],
[trDat[i][0][1] for i in range(len(trDat))],
c=[trDat[i][1] for i in range(len(trDat))],
cmap=classColormap)
plt.xlabel("Кількість негативних відгуків ")
plt.ylabel("Тривалість дільності, (Роки)")
plt.grid()
plt.show()

```

Таким чином, можна виділити категорії постачальників за окремими класами, класифікаційні ознаки яких встановлено з використанням методу k -найближчих сусідів (рис. 3).

Так видно, що класифікація, яка реалізована на основі даного може бути застосована в процесі інформаційно-аналітичного забезпечення економічної безпеки молокопереробних підприємств.

Аналіз даного алгоритму на предмет помилок та достовірності не проводився, що вимагає подальших досліджень.

Висновки. Метод k -найближчих сусідів дозволяє з високою ймовірністю визначати надійних постачальників сировини для молокопереробних підприємств, але при цьому точність виявлення таких прогнозів потребує подальшого дослідження. Навіть при великих обсягах контрольної вибірки даних, потребує подальших досліджень процес встановлення кількості найближчих точок, встановлення оптимального алгоритму визначення відстані між точками, а також програмні засоби, оскільки обробка великих масивів даних вимагає значних обчислювальних потужностей.

Список використаних джерел:

1. Седов А.В. Закупки: причины, классификация, типовые решения. *Вестник Ульяновского государственного технического университета* 2017. № 3 (79). С. 65–69.
2. Рахматуллина В.Р., Горшенин В.Ф. (2018). Цифровая трансформация закупочной логистики. *Общество, экономика, управление*. 2018. № 4. С. 40–45.
3. Gonçalo Thomas Edson Espíndola, Alencar Luciana Hazin. A supplier selection model based on classifying its strategic impact for a company's business results. *Pesquisa Operacional*. 2014. № 34(2). P. 347–369.
4. Батура Т.В. Методы автоматической классификации текстов. *Программные продукты и системы*. 2017. № 30 (1). С. 85–89. DOI: 10.15827/0236-235X.030.1.085-099.
5. Классификатор kNN. IT спільнота Хабр. 2020. URL: <https://habr.com/ru/post/149693>.
6. Попело В.Д., Ванеева М.В. Теория математической обработки геодезических измерений. Часть 2. Оценивание результатов геодезических измерений и их погрешностей на основе вероятностных представлений: учебное пособие. Воронеж : ВГАУ, 2015. 138 с.
7. Sandro Tosi. *Matplotlib for Python Developers*. Packt Publishing, 2009. 308 p.

References:

1. Sedov A.V. (2017) Zakupki: prichinyi, klassifikatsiya, tipovyye resheniya. *Vestnik Ulyanovskogo gosudarstvennogo tehniceskogo universiteta*, no. 3 (79), pp. 65–69.
2. Rahmatullina V.R., Gorshenin V.F. (2018). Tsifrovaya transformatsiya zakupochnoy logistiki. *Obschestvo, ekonomika, upravlenie*, no. 4, pp. 40–45.
3. Gonçalo Thomas Edson Espíndola, Alencar, Luciana Hazin (2014) A supplier selection model based on classifying its strategic impact for a company's business results. *Pesquisa Operacional*, no. 34(2), pp. 347–369.
4. Batura T.V. (2017). Metodyi avtomaticheskoy klassifikatsii tekstov. *Programmnyie produktyi i sistemyi*, no. 30 (1), pp. 85–89. DOI: 10.15827/0236-235X.030.1.085-099.
5. Klassifikator kNN. IT spIlnota Habr. 2020. Available at: <https://habr.com/ru/post/149693>.
6. Popelo V.D., Vaneeva M.V. (2015) Teoriya matematicheskoy obrabotki geodezicheskikh izmereniy. Chast 2. Otsenivanie rezultatov geodezicheskikh izmereniy i ih pogreshnostey na osnove veroyatnostnyih predstavleniy: uchebnoe posobie. Voronezh: VGAU, 138 p.
7. Sandro Tosi (2009) *Matplotlib for Python Developers*. Packt Publishing, 308 p.