



original article | UDC 004.4'2: 631.526.3 | doi: 10.31210/visnyk2019.04.30

PECULIARITIES OF USING MACHINE LEARNING TOOLS DURING IDENTIFICATION OF SIMILAR PLANT VARIETIES (*Lactuca sativa* L. var. *capitata* as example)

N. S. Orlenko,

ORCID ID: [0000-0003-4103-7806](https://orcid.org/0000-0003-4103-7806), E-mail: sops@sops.gov.ua,

N. V. Leschuk,

ORCID ID: [0000-0001-6025-3702](https://orcid.org/0000-0001-6025-3702), E-mail: nadiya1511@ukr.net,

N. V. Symonenko,

M. M. Tagantsova,

O. A. Stadnichenko,

Ukrainian Institute of Plant Variety Expert Examination, 15, Henerala Rodymtseva str., Kyiv, 03041, Ukraine

*Technological approaches to using Machine Learning tools in the identification of new crop varieties based on the data set of well-known varieties, by using the algorithm of the nearest neighbor were considered on the example of morphological description data of *Lactuca sativa* L. var. *capitata* varieties. The aim of the work is to give an example of Machine Learning method and assess the suitability of its using while processing data on codes of plant varieties' morphological description. To substantiate the suitability of using machine learning in processing the results of examining difference, uniformity and stability, in particular the identification of groups of similar varieties of plants analytical and statistical methods were used. The information technology of using machine learning tools for the formation of a computer model of similar lettuce varieties by applying the IBM SPSS Statistics statistical package has been developed. The results of the experiment with a computer-based training model revealed that the most accurate classification results were obtained by using the sign of the size of lettuce head as the target variable model and the sign of lettuce head density as the focal variable. The study showed the suitability of using the Machine Learning tool to identify groups of similar lettuce varieties by morphological features. The IBM SPSS Statistics package of statistical programs is convenient in using, it provides the researcher with a wide range of tools for experimenting with a botanical taxon variety model, and enables to visualize the obtained modeling results using diagrams that clearly show the results. The main diagram of the model is interactive, allowing the researcher to experiment with the model. This method can be recommended for using in processing qualification expertise data for differences, uniformity and stability.*

Key words: expertise of difference, uniformity and stability, Crisphead lettuce, statistical methods in breeding, Machine Learning, IBM SPSS Statistics.

ОСОБЛИВОСТІ ВИКОРИСТАННЯ ЗАСОБІВ MACHINE LEARNING ПІД ЧАС ІДЕНТИФІКАЦІЇ ПОДІБНИХ СОРТІВ РОСЛИН (на прикладі *Lactuca sativa* L. var. *capitata*)

Н. С. Орленко, Н. В. Лещук, Н. В. Симоненко, М. М. Таганцова, О. А. Стадніченко,

Український інститут експертизи сортів рослин, м. Київ, Україна

Розглянуто технологічні підходи до використання засобів Machine Learning під час ідентифікації нових сортів сільськогосподарських рослин на підставі набору даних загальновідомих сортів, алгоритм найближчого сусіда. Метою цієї роботи є розроблення прикладу застосування методу машинного навчання та оцінювання придатності його використання під час оброблення даних прояву морфологічного опису сортів рослин. Під час дослідження використано й аналітичний та статис-

тичний методи. Дослідження проводилось на прикладі даних морфологічного опису сортів *Lactuca sativa L. var. capitata*. Опробована інформаційна технологія використання засобів машинного навчання для формування комп'ютерної моделі подібних сортів салату посівного головчастого з використанням статистичного пакету IBM SPSS Statistics. У результаті експерименту з комп'ютерною моделлю навчання встановлено, що найбільш точні результати класифікації отримано з використанням ознаки розмір головки салату посівного головчастого як цільової змінної моделі та ознаки щільність головки салату посівного як фокусну змінну. Дослідження показало придатність застосування засобу Machine Learning під час ідентифікації груп подібних сортів салату посівного головчастого за морфологічними ознаками. Пакет статистичних програм IBM SPSS Statistics є зручним у користуванні, надає досліднику широкий спектр засобів експериментування з моделлю сортів ботаничного таксону, дає змогу візуалізувати отримані результати моделювання з використанням діаграм, які добре унаочнюють результати моделювання. Головна діаграма моделі є інтерактивною, що дозволяє досліднику експериментувати з моделлю. Цей метод може бути рекомендовано для використання під час оброблення даних кваліфікаційної експертизи на відмінність, однорідність та стабільність.

Ключові слова: експертиза на відмінність, однорідність та стабільність, салат посівний головчастий, статистичні методи в селекції, Machine Learning, IBM SPSS Statistics.

ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ СРЕДСТВ MACHINE LEARNING ПРИ ИДЕНТИФИКАЦИИ ПОХОЖИХ СОРТОВ РАСТЕНИЙ (на примере *Lactuca sativa L. var. capitata*)

Н. С. Орленко, Н. В. Лещук, Н. В. Симоненко, М. Н. Таганцова, О. А. Стадниченко,
Украинский институт экспертизы сортов растений, г. Киев, Украина

Рассмотрены технологические подходы к использованию средств Machine Learning при идентификации новых сортов сельскохозяйственных растений на основании набора данных общеизвестных сортов. В ходе исследования использованы аналитический, статистический методы, алгоритм ближайшего соседа. Исследование проводилось на примере данных морфологического описания сортов *Lactuca sativa L. var. capitata*. Апробирована информационная технология использования средств машинного обучения для формирования компьютерной модели подобных сортов салата посевого кочанного с использованием статистического пакета IBM SPSS Statistics. В результате эксперимента с компьютерной моделью установлено, что наиболее точные результаты классификации получены с использованием признака размер кочана салата посевого кочанного в качестве целевой переменной модели и признака плотность кочана салата посевого в качестве фокусной переменной. Исследование показало пригодность применения средства Machine Learning при идентификации групп подобных сортов салата посевого головчастого по морфологическим признакам. Пакет статистических программ IBM SPSS Statistics является удобным в пользовании, предоставляет исследователю широкий спектр средств экспериментирования с моделью сортов ботанического таксона, позволяет визуализировать полученные результаты моделирования с использованием диаграмм, которые хорошо демонстрируют результаты моделирования. Главная диаграмма модели является интерактивной, что позволяет исследователю экспериментировать с моделью. Этот метод может быть рекомендован для использования при обработке данных квалификационной экспертизы на отличие, однородность и стабильность.

Ключевые слова: экспертиза на отличие, однородность и стабильность, салат посевной кочанный, статистические методы в селекции, Machine Learning, IBM SPSS Statistics

Вступ

Відомо, що сорт є основною ланкою технології вирощування сільськогосподарських культур. Для точної ідентифікації груп подібних сортів, важливим питанням є правильний обраний математико-статистичний апарат класифікації сортів рослин за морфологічними ознаками. Більшість з опублікованих досліджень, які присвячені ідентифікації подібних сортів сільськогосподарських рослин, висвітлюють питання використання кластерного аналізу під час ідентифікації, як приклад роботи [11–13]. В роботі [15] проводиться аналіз придатних для використання статистичних методів під час оброблення

даних щодо тканин рослин й також проаналізовано метод ієрархічної класифікації. Докладно вплив на результат різних ієрархічних агрегативних методів і метрик, таких як методи: ближнього сусіда, дальнього сусіда, середнього зв'язку, середнього сусіда та метода Варда розглянуто в роботі [8].

Предметом дослідження є інформаційна технологія застосування методу машинного навчання під час ідентифікації подібних сортів рослин в процесі оброблення даних експертизи на відмінність, однорідність та стабільність.

Назву «машинне навчання» (Machine learning) було започатковано 1959 року Артуром Семюелем [21]. Цей метод вже широко використовується у сфері біоінформатики та класифікація послідовностей ДНК.

Об'єктом цього дослідження є дані морфологічного опису сортів ботанічного таксону салату посівного головчастого. Салат посівний (*Lactuca sativa* L.) – найважливіша овочева рослина в групі зеленних. Щорічно селекціонерами створюється велика кількість сортів салату посівного різних типів, що характеризується значними морфологічними відмінностями [4] та включає сім основних груп культурних сортів, що відрізняється фенотипічно [15, 16, 18]. Розведення салату первинно орієнтовано на різні морфологічні особливості та стійкість до коливання температур, стійкість проти хвороби та шкідників [1, 10, 15], тому важливо серед кожної групи салатів ідентифікувати схожі сорти кожної групи. В межах цього дослідження було обрано ботанічний таксон *Lactuca sativa* L. var. *capitata*. Отримана група представлена 38 сортами вітчизняного та іноземного походження (Україна, Нідерланди, Німеччина, Франція, Чеська Республіка у Державному реєстрі сортів рослин, придатних для поширення в Україні) станом на вересень 2019 р. [2].

Особливістю даного дослідження є те, що запропонований метод ідентифікації, який відноситься до групи методів інтелектуального аналізу даних Machine Learning [3] дозволяє використовувати загально відому колекцію сортів, що знаходиться в БД ІАС Українського інституту експертизи сортів рослин в якості підґрунтя для ідентифікації нових сортів.

Аналіз являє собою метод класифікації спостережень, які отримані в результаті експертизи на однорідність, відмінність та стабільність (ВОС) на основі подібності отриманих результатів дослідження морфологічних ознак сортів рослин (спостережень за морфологічними ознаками на різних фенологічних стадіях росту рослин та даним лабораторного аналізу прояву властивостей рослин). Цей метод ще має назву метод машинного навчання [6] і був розроблений в якості способу розпізнавання структури даних, що мають не точну відповідність та складну структуру спостережень. Подібні спостереження – це спостереження, що близькі один до одного, а не схожі спостереження, навпаки, віддалені один від одного. Таким чином, дистанція між двома спостереженнями є критерієм їх відмінності.

Близькими один до одного спостереження називаються «сусіди» [13]. Коли з'являється нове спостереження, воно позначається знаком питання, а потім обчислюється відстань кількісних, якісних та псевдоякісних значень його ознак (у нашому випадку набору морфологічних характеристик сортів рослин) від всіх інших спостережень в моделі ботанічного таксону рослин. Пакет прикладних статистичних програм IBM SPSS Statistics автоматично проводить класифікацію найбільш схожих спостережень і нове спостереження розміщується в категорію, в якій є найбільша кількість спостережень з найближче схожими ознаками [7, 20].

Метою даної роботи є надання прикладу застосування методу машинного навчання та оцінювання придатності його застосування під час оброблення даних кодів прояву морфологічного опису сортів рослин за експертизи на ВОС. Для досягнення цієї мети було поставлено такі завдання: – визначити вплив показників змінних обраних в якості цільової та фокусної під час формування комп'ютерної моделі на результат класифікації сортів рослин салату головчастого; – проаналізувати юзабіліті [18] застосування цього методу в пакеті IBM SPSS Statistics.

Матеріали і методи досліджень

Під час дослідження використано дані результатів випробувань сортів салату посівного головчастого на відмінність, однорідність і стабільність (ВОС) (tests of a selection achievement for distinctiveness, uniformity and stability) за період 1996–2018 рр., які проводилися Українським інститутом експертизи сортів рослин відповідно до міжнародних вимог і відповідно до Guidelines for the conduct of tests for distinctness, uniformity and stability TG/13/11 LETTUCE. UPOV Code(s): LACTU_SAT [22] та методики проведення експертизи сортів салату посівного [5].

Оброблення результатів експертизи проводились з використанням методу машинного навчання зі

застосуванням алгоритму найближчих сусідів. Згідно з цим алгоритмом, передбачається, що вже є якась кількість об'єктів з точною класифікацією (у нашому випадку – подібних сортів рослин салату посівного головчастого), і потрібно виробити правило, яке дозволяє віднести новий сорт до одного з можливих класів (набору подібних за морфологічними ознаками сортів рослин).-*KNN* алгоритм здійснює підбір коефіцієнтів, які визначають міру подібності, для нових сортів салату головчастого, а *k* – це кількість записів, які будуть вважатися близькими з використанням таких правил:

- $(x,y) \geq 0, d(x,y) = 0$ тоді і лише тоді коли $x = y$;
1. $d(x,y) = d(y,x)$;
 2. $d(x,z) \leq d(x,y) + d(y,z)$, за умови, що точки x, y, z не лежать на одній прямій.

Де x, y, z – це вектори ознак, об'єктів які порівнюються. Впорядкування значень атрибутів проводиться з використанням відстані Евкліда.

У якості цільової змінної моделі (target optional) обрано ознаку «Головка: розмір», фокусна змінна (focal case identifier) – «Головка: за щільністю», мітка спостережень (case label) – «назва сорту». Перелік показників (features) складають такі ознаки: «Насіннина: забарвлення», «Сіянець: антоціанове забарвлення», «Сіянець: сім'ядолі за розміром (за повного розвитку)», «Листок: положення у стадії 10-12 листків», «Листкова пластинка: за розсіченістю краю», «Рослина: діаметр», «Рослина: утворення головки», «Головка: за щільністю», «Головка: розмір», «Листок: за товщиною», «Листок: положення за збиральної стиглості», «Листок: форма», «Листок: форма верхівки», «Листок: інтенсивність забарвлення зовнішніх листків».

Для побудови моделі групи сортів в параметрі «кількість найближчих сусідів (*k*) задано визначити три найбільш схожих сортів. В якості альтернативи (forced entry) задано показник «Рослина: діаметр», що дозволяє процедурі IBM SPSS Statistics вибрати оптимальне число сусідів.

Результати досліджень та їх обговорення

У результаті моделювання було сформовано модель подібних сортів *Lactuca sativa* L. var. Capitata, яка включала 30 сортів рослин, які були використані у навчальній вибірці та 12 сортів салату головчастого у контрольній групі. Тобто 71,4 % від загальної кількості сортів, що брали участь у розрахунках, було віднесено до навчальної вибірки, а 28,6 % до контрольної групи. Результатні дані візуалізовано на п'яти діаграмах Моделі найбільшої подібності. Головна діаграма моделі «Predictor Space» (Простір показників) унаочнює оглядове зображення моделі. Діаграма простору показників є інтерактивною. Кожна вісь представляє показник в моделі, а розташування точок на діаграмі показує значення цих показників для спостережень в навчальній і контрольній групах. Ця діаграма демонструє зв'язок між даними навчальної та результатної вибірки що є «найближчими сусідами». Зауважимо, що колом позначено сорти, які входять до складу навчальної вибірки, а трикутником позначені сорти контрольної вибірки. До навчальної вибірки увійшли сорти 'Авірам', 'Амадеус', 'Бакеро', 'Бернардінас', 'Вайдоза', 'Джиска', 'Драгон', 'Імеджінейшн', 'Кірен', 'Кірібаті', 'Кірінія', 'Кісмі', 'Крунчита', 'Міретт', 'Нанетт', 'Нейшн', 'Ноблес', 'Отілі', 'Платінас', 'Рекорд', 'Руксай', 'Сантарінас', 'Фріліліс', 'Паджеро', 'Дивограй', 'Рубетт', 'Гондар', 'Аргентінас', 'Сільвінас' та 'Діамантінас'. Елемент діаграми «Target» засвідчує, яку змінну обрано як цільову.

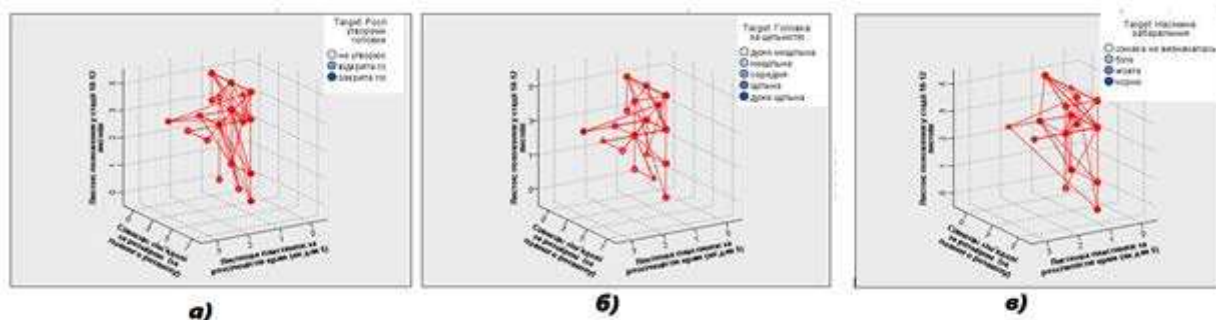


Рис. 1. Фрагменти головних діаграм моделей найбільшої подібності «Predictor Space» (Простір показників)

Експеримент з моделлю схожих сортів рослин показав залежність результату від обраної цільової змінної. Моделювання проводилось з використанням в якості цільових змінних ознак «Рослина: утворення головки» (рис. 1 а), «Головка за щільністю» (рис. 1 б), «Насінина: забарвлення» (рис. 1 в), «Розмір головки рослини» «Головка: розмір» Встановлено, що найкращі результати групування подібних сортів отримано за використанням в якості цільової змінної ознаки «Головка: розмір». Мітка Build model selected peredicat, засвідчує той факт, що модель була побудована для ідентифікації трьох подібних сортів для кожного сорту салату головчастого.

Колір точки показує значення цільової змінної для даного спостереження. Різними кольорами позначається приналежність до різних категорій категоріальної цільової змінної, а саме фокусних та не фокусних змінних. Більш жирний контур вказує на те, що спостереження є фокусною. Фокусні спостереження показуються з'єднаними з їх k найближчими сусідами. Виміри діаграми «Листкова пластинка: за розсіченістю краю верхівки», «Сіянець: сім'ядолі за розміром (за повного розвитку)» та «Листок: положення за збиральної стиглості (зовнішні листки у головчастого салату або сформовані листки для листкового і стеблового салатів)» обрані пакетом прикладних програм автоматично.

На рисунку 2 представлено діаграму важливості параметрів, яка дозволяє сконцентрувати увагу на найбільш важливих ознаках морфологічного прояву сортів салату головчастого під час побудови моделі, та відмітити менш суттєві ознаки.

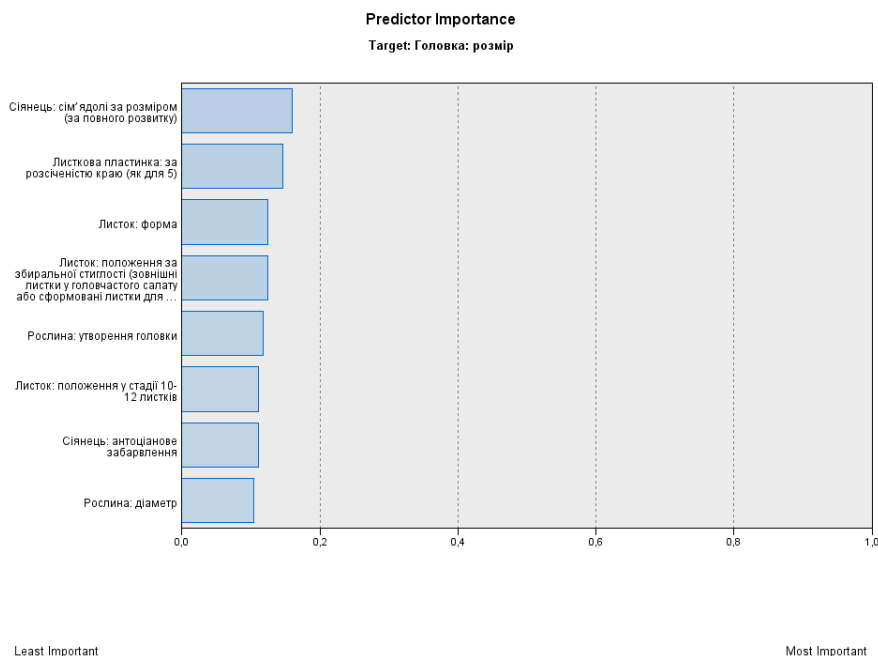


Рис. 2. Діаграму важливості параметрів

Приклад діаграми квадрантів (Peer Chart) наведено на рисунку 3. Ця діаграма унаочнює виведення фокусних спостережень та їх k найближчих сусідів на діаграмі. Діаграма розбита на панелі за показниками.



Рис. 3. Фрагменти діаграми квадрантів (Peer Chart) для ознак

ТЕХНІЧНІ НАУКИ

Повний перелік груп, виокремлено за тринадцяти морфологічними ознаками, сортів салату наведено у таблиці. Ця таблиця відображає три найближчих сусідів до фокусного сорту, а також значення відстані до них. Стовпчики таблиці містять назву сорту фокусного спостереження, назви трьох найбільш подібних сортів, а також значення відстаней.

Групування подібних сортів за відстанню до найближчих сусідів (k Nearest Neighbor and Distance)

Назва сорту базового (фокусного) сорту	Назва першого подібного сорту	Назва другого подібного сорту	Назва третього подібного сорту	Відстань у порівнянні з 1-им	Відстань у порівнянні з 2-им	Відстань у порівнянні з 3-м
Дивограй	Сільвінас	Діамантінас	Бернардінас	0,525	0,525	0,525
Перл Джем	Амадеус	Куала	Аквіно	0,788	0,987	0,987
Амадеус	Аквіно	Перл Джем	Куала	0,750	0,788	0,987
Картагенас	Лагунас	Міретт	Куала	0	0,275	0,487
Платінас	Картагенас	Лагунас	Міретт	0,564	0,564	0,564
Нанетт	Фіоретт	Міретт	Аргентінас	0	0,462	0,5
Афіціон	Кісмі	Джиска	Бакеро	1,013	1,038	1,186
Куала	Картагенас	Лагунас	Аквіно	0,487	0,487	0,725
Галера	Сільвінас	Діамантінас	Бернардінас	0,288	0,288	0,288
Фрілліс	Авірам	Імеджінейшн	Драгон	0,288	0,288	0,487
Аргентінас	Сільвінас	Діамантінас	Бернардінас	0,199	0,199	0,199
Сільвінас	Діамантінас	Бернардінас	Аргентінас	0	0	0,199
Діамантінас	Сільвінас	Бернардінас	Аргентінас	0	0	0,199
Бернардінас	Сільвінас	Діамантінас	Аргентінас	0	0	0,199
Джиска	Авірам	Міретт	Кісмі	0,674	0,775	0,788
Імеджінейшн	Авірам	Фрілліс	Драгон	0,288	0,288	0,487
Нейшн	Кірінія	Кітонія	Авірам	0,674	0,725	0,763
Руксай	Кітонія	Нейшн	Фрілліс	0,275	1	1,186
Кітонія	Руксай	Нейшн	Фрілліс	0,275	0,725	0,911
Аквіно	Куала	Картагенас	Лагунас	0,725	0,725	0,725
Кісмі	Галера	Аргентінас	Сільвінас	0,775	0,775	0,775
Патішн	Джиска	Авірам	Перл Джем	0,987	1,025	1,115
Кірінія	Нейшн	Авірам	Драгон	0,674	0,75	0,75
Кірібаті	Фрілліс	Кірінія	Авірам	0,775	1	1,064
Кірен	Аквіно	Куала	Кісмі	0,814	0,962	1,064
Авірам	Фрілліс	Імеджінейшн	Драгон	0,288	0,288	0,487
Вайдоза	Кафтан	Айс Вейв	Імеджінейшн	0,212	0,436	0,712
Бакеро	Вайдоза	Платінас	Кафтан	0,75	0,788	0,962
Кафтан	Вайдоза	Айс Вейв	Імеджінейшн	0,212	0,648	0,924
Айс Вейв	Вайдоза	Кафтан	Імеджінейшн	0,436	0,648	0,712
Айс Сіркл	Міретт	Аргентінас	Сільвінас	0,962	0,987	0,987
Драгон	Авірам	Фрілліс	Імеджінейшн	0,487	0,487	0,487
Вісмар	Імеджінейшн	Грін мун	Авірам	0,564	0,763	0,852
Грін мун	Драгон	Вісмар	Авірам	0,576	0,763	0,775

Аналіз за діапазонами відстаней даних таблиці свідчить, що у діапазоні відстаней від 0 до 0,5 виокремлено дванадцять груп сортів, які мають три подібні до базового сорти в цьому діапазоні значення коефіцієнти. У діапазоні від 0,5001 до 0,999 виокремлено одинадцять груп сортів, а коефіцієнт більший за одиницю розрахована для трьох груп сортів.

Зауважимо, що нульові коефіцієнти відстаней, які розраховано для сортів 'Аргентінас', 'Бернардінас', 'Діамантінас' та 'Сільвінас' свідчать про те, що ці сорти є подібними за 13 ознаками, які відібрані з 42 ознак визначених методикою проведення експертизи на ВОС. І це засвідчує, що сорти є найбільш подібними сортами, якщо їх порівнювати за тринадцятьма виокремленими зі списку, який складає 42 ознаки, що рекомендовано методикою. Однак, кожний з цих сортів, має хоча б одну відмінну ознаку: сорт 'Аргентінас' має більш щільну головку, сорт 'Бернардінас' має меншу головку за розміром, сорт 'Діамантінас' має менший діаметр рослини, а сорт 'Сільвінас' відрізняється забарвленням зовнішніх листків.

Сорти 'Руксай' та 'Кітонія' є подібними між собою, але мають значну відстань від сортів 'Нейши' та 'Фрілліс'.

Висновки

Встановлено, що машинне навчання досить ефективним засобом аналізу даних експертизи на ВОС та значно полегшує пошук закономірностей серед великого набору даних. На прикладі ботанічного таксону салату посівного головчастого виокремлено групи найбільш схожих сортів. За 13-ю морфологічними ознаками виокремлено такі групи сортів найбільш подібних сортів; перша група – 'Сільвінас', 'Діамантінас', 'Бернардінас' та 'Аргентінас', друга група – 'Картагенас', 'Лагунас', 'Міретт' та 'Куала'; третя група – 'Нанетт', 'Фіоретт', 'Міретт' та 'Аргентінас'; четверта група – 'Галера', 'Сільвінас', 'Діамантінас' та 'Бернардінас'. Розмежування вихідних даних на навчальні та контрольні, дозволяє провести «тренування» моделі на наборі даних загально відомих сортів салату. Експериментальним шляхом встановлено, що найбільш адекватна модель груп подібних сортів салату посівного головчастого формується у разі використання ознаки «Головка: розмір» як цільової змінної та ознаки «Головка: за щільністю» як фокусної змінної. Отримані результати свідчать про перспективність використання алгоритму найближчих сусідів під час ідентифікації подібних сортів рослин.

Перспективи подальших досліджень. У базі даних Українського інституту експертизи сортів рослин зберігаються дані щодо 46 303 сортів рослин 661 ботанічного таксону. Дослідження показало придатність використання засобів машинного навчання під час ідентифікації подібних сортів салату посівного головчастого. Використання цього методу поширено під час оброблення результатів експертизи на відмінність, однорідність та стабільність сортів рослин інших ботанічних таксонів.

References

1. Grinko, N. N. (2011). *Vospriimchivost k virusu zheltoj mozaiki kochannyh sortov salata. Zashita i Karantin Rastenij*, 4, 33–34 [In Russian].
2. Derzhavnyi reestr sortiv roslyn, prydatnykh dlia poshyrennia v Ukraini na 2019 rik. (2019). Retrieved from: <https://sops.gov.ua/reestr-sortiv-roslyn> [In Ukrainian].
3. Leskovets, Yu., Radzharaman, Yu., & Ulman, D. (2016). *Analiz bol'shikh naborov dannykh*. Moscow: DMK [in Russian].
4. Leshchuk, N. V., Kryvytskyi, K. M., Maister, N. V., Bronovytska, M. A. (2010). *Kompleksna otsinka biolohichnoho potentsialu sortovykh resursiv salatu posivnoho (Lactuca sativa L.). Sortovyvchennia ta Okhorona Prav na Sorty Roslyn*, 2, 63–70 [In Ukrainian].
5. Leshchuk, N. V. (Red.). (2014). *Metodyka provedennia ekspertyzy sortiv roslyn hrupy ovochevykh, kartopli ta hrybiv na vidminnist, odnoridnist i stabilnist. Ukrainskyi instytut ekspertyzy sortiv roslyn*. Retrieved from: <https://sops.gov.ua/vos> [In Ukrainian].
6. Marmanis, Kh., & Babenko, D. (2011). *Algoritmy intelektual'nogo Interneta. Peredovye metodiki sbora, analiza i obrabotki dannykh*. Moscow: Simvol [In Russian].
7. Nasledov, A. D. (2013). *IMB SPSS Statistics 20 i AMOS: professional'nyy statisticheskiiy analiz dannykh*. St. Petersburg: Piter [In Russian].
8. Orlenko, N. S., Mazhuha, K. M., Dushar, M. B., & Maslechkin, V. V. (2019). *porivnialnyi analiz iierarkhichnykh metodiv klasterizatsii, prydatnykh dlia obroblennia danykh morfolohichnykh oznak sortiv*

roslyn. *Visnyk Poltavskoi derzhavnoi ahrarnoi akademii*, (2), 261–269. doi: 10.31210/visnyk2019.02.35 [In Ukrainian].

9. Osipova, G. S., Kondrat'ev, V. M., & Yakovleva, M. G. (2015). Agrobiological assessment of lettuce and half-lettuce varieties in the autumn turnover of film greenhouses in the Leningrad Region. *Nauchnyy vklad molodykh issledovateley v sokhranenie traditsiy i razvitie APK: sb. Mezhdunar. nauchno-prakt. konf. molodykh uchenykh i studentov*. St. Petersburg, Russia [In Russian].

10. Tyshchenko, V. M. (2005). Cluster analysis as a method of individual selection of high-yield winter wheat plants in F₂. *Selektsiia i Nasinnytstvo*, 89, 125–137 [In Ukrainian].

11. Tishchenko, V. N., Chekalin, N. M., & Zyukov, M. E. (2004). Using cluster analysis to identify and select highly productive winter wheat genotypes in the early stages of selection. *Fakt. Eksp. Evol. Org.*, 2, 270–278 [In Russian]

12. Tishchenko, V. N. (2004). Effektivnost ispolzovaniya novogo selekcionnogo indeksa v selektsii ozimoy pshenicy. *Fakty Eksperimentalnoi Evoliutsii Orhanizmiv* 2, 266–270 [In Russian].

13. Lantz, B. (2013). *Machine Learning with R*. Birmongham-Mumbai: Pack Publishing.

14. Compton, M. E. (1994). Statistical methods suitable for the analysis of plant tissue culture data. *Plant Cell Tiss Organ Cult*, 37 (3), 217–242. doi: 10.1007/BF00042336.

15. Craker, L. E., & Seibert, M. (1982). Light Energy Requirements for Controlled Environment Growth of Lettuce and Radish. *Transactions of the ASAE*, 25 (1), 0214–0216. doi: 10.13031/2013.33506.

16. Dufault, R. J., Ward, B., & Hassell, R. L. (2006). Planting Date and Romaine Lettuce Cultivar Affect Quality and Productivity. *HortScience*, 41 (3), 640–645. doi: 10.21273/hortsci.41.3.640

17. Helm, J. (1954). *Lactuca sativa* L. in morphologisch-systematischer Sicht. *Die Kulturpflanze*, 2 (1), 72–129. doi: 10.1007/BF02095730.

18. Nielsen, J. (1994). Usability inspection methods. *Conference Companion, CHI'940*. Boston: Massachusetts USA.

19. Kalloo, & Krug, H. (1980). Sortendifferenzierung bei Kopfsalat (*Lactuca sativa* var. *capitata*) – Vorläufige Mitteilung. *Die Gartenbauwissenschaft*, 451 (3), 101–103.

20. Levesque R. (2007). *SPSS programming and data management: a guide for SPSS and SAS users: 4th ed*. Chicago Ill: SPSS Inc.

21. Samuel, A. L. (1988). *Some Studies in Machine Learning Using the Game of Checkers. I. Computer Games I*. New York: Springer. doi: 10.1007/978-1-4613-8716-9_14.

22. Test Guidelines for the conduct of tests for distinctness, uniformity and stability of Lettuce (*Lactuca sativa* L.) (TG/13/10 Rev. 2, UPOV). Geneva. Retrieved from: www.upov.int/edocs/tgdocs/en/tg013.pdf.

Стаття надійшла до редакції 08.10.2019 р.

Бібліографічний опис для цитування:

Орленко Н. С., Лецук Н. В., Симоненко Н. В., Таганцова М. М., Стадніченко О. А. Особливості використання засобів machine learning під час ідентифікації подібних сортів рослин (на прикладі *Lactuca sativa* L. var. *capitata*). *Вісник ПДАА*. 2019. № 4. С. 233–240.

© Орленко Наталія Станіславівна, Лецук Надія Василівна, Симоненко Наталія Володимирівна, Таганцова Марина Миколаївна, Стадніченко Ольга Анатоліївна, 2019