

КЛАСТЕРИЗАЦИЯ ОРГАНОЛЕПТИЧЕСКИХ ОЦЕНОК КАЧЕСТВА КРАСНЫХ И БЕЛЫХ ВИН ПО ФИЗИКО-ХИМИЧЕСКИМ ПОКАЗАТЕЛЯМ С ПОМОЩЬЮ ПЛАТФОРМЫ STATISTICA

© 2023 г. Н. В. Бондарев^{1,*}

¹ Харьковский национальный университет имени В. Н. Каразина, пл. Свободы 4, Харьков, 61022 Украина

*e-mail: n_bondarev@ukr.net

Поступило в редакцию 15 августа 2023 г.

После доработки 15 августа 2023 г.

Принято к печати 16 августа 2023 г.

С помощью компьютерного анализа выполнена кластеризация органолептических оценок по десяти-балльной шкале экспертов по вину и экспериментальных физико-химических показателей красных (1599 образцов) и белых (4898 образцов) вин португальских производителей. Методами агломеративной и итеративной (алгоритм *k*-средних) кластеризации выявлено группирование сходных образцов вин в три, четыре и шесть кластеров в зависимости от евклидова расстояния объединения. Установлена количественная наполняемость кластеров образцами плохих (оценки 3 и 4), нормальных (оценки 5 и 6) и хороших (оценки 7, 8, 9) вин. Выполнены нейросетевой (MLP) и дискриминантный (DA) анализ; задействованы алгоритмы деревьев классификации (СТ), машины опорных векторов (SVM), наивной баесовской классификации (NBC) и ближайших соседей (kNN). Лучшую производительность продемонстрировали нейросетевые модели. Обучены многослойные персептроны-классификаторы: для красных вин – MLP 11-7-3, MLP 11-13-4, MLP 11-14-6; для белых вин – MLP 11-9-3, MLP 11-5-4, MLP 11-9-6. Выявлены свойства вин, вклад которых в разделяющую мощь классификаторов является определяющим. Приведены диапазоны изменения физико-химических показателей в трех кластерах красных и белых вин для плохих, нормальных и хороших вин.

Ключевые слова: вино, качество, кластеризация, классификация, прогнозирование, STATISTICA

DOI: 10.31857/S0044460X23110161, **EDN:** PCZYWD

Качество вина, как и любого пищевого продукта, влияет на здоровье человека. Поэтому проблемы качества винодельческой продукции постоянно находятся в поле зрения представителей различных областей знания. В вине обнаружено более 350 химических веществ [1]. В этой связи уместно привести изречение Парацельса (1493–1541) – швейцарского врача, алхимика, философа, естествоиспытателя, натурфилософа эпохи Возрождения: «Все есть яд, и ничто не лишено ядовитости; одна лишь доза делает яд незаметным»

(«Alle Dinge sind Gift, und nichts ist ohne Gift, allein die Dosis macht dass ein Ding kein Gift ist» [2]).

Выявление главных факторов качества вина часто проводится на основе органолептических оценок экспертов по вину и физико-химических лабораторных тестов. Способам и методам установления взаимосвязи субъективных винных предпочтений экспертов и объективных физико-химических показателей вина посвящены работы носителей универсальных компетенций [3–22].

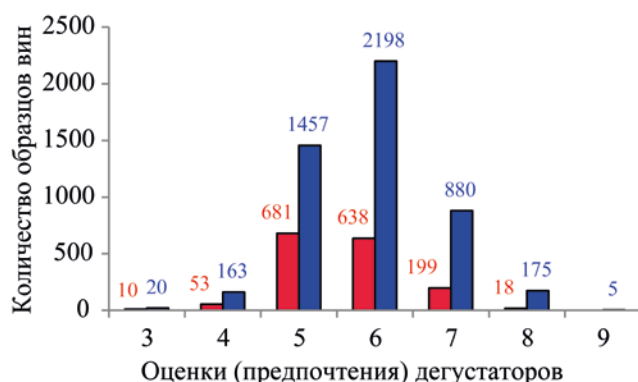


Рис. 1. Диаграмма распределения количества образцов красных и белых (столбцы окрашены в синий цвет) вин по оценкам (предпочтениям) экспертов.

Цель данной работы заключается в кластеризации органолептических предпочтений экспертов по вину на основе физико-химических показателей образцов вин на платформе STATISTICA. Экспериментальные данные о качестве вина взяты из общедоступного для исследовательских целей репозитория машинного обучения Калифорнийского университета в Ирвине (UCI) [23]. Два набора данных содержат 1599 образцов красного и 4898 образцов белого вина с 11 физико-химическими показателями для каждого образца. Вина (Vinho Verde) были приготовлены португальскими производителями из винограда, собранного с мая 2004 г. по февраль 2007 г. в северо-западном регионе Минью (Minho) Португалии. Входные данные – это объективные тесты, выходные – сенсорные оценки, выставленные по десятибалльной шкале экспертами по вину. Физико-химические лабораторные показатели вин включают фиксированную кислотность (FA), летучую кислотность (VA), лимонную кислоту (CA), остаточный сахар (RS), хлориды (Cl), свободный диоксид серы (FDS), общий диоксид серы (TDS), плотность (D), pH, сульфаты (SU) и спирт (Al). Информация о сортах винограда, марке вина и т. п. не приведена. Классы вин упорядочены, но не сбалансированы – нормальных вин гораздо больше, чем отличных и плохих [7] (рис. 1).

Принимая во внимание крайнюю неравномерность распределения образцов вин по оценкам

экспертов (рис. 1), в работе предлагается такая градация качества вин: 3, 4 – плохое вино; 5, 6 – нормальное вино, 7–9 – хорошее вино на основе 10-балльной шкалы (3–9), приведенной в работах [7, 23]. Следует отметить, что градация может быть и иной, вплоть до исключения из компьютерного анализа малочисленных образцов вин с низкими и высокими оценками [22]. Это зависит от того, какие цели определены исследователем-руководителем своим последователям в рамках использования общедоступных данных для машинного обучения методами бурно развивающихся современных информационных технологий.

Описание физико-химических показателей вин [1]. Кислотность – один из основных показателей химического состава и вкусовых признаков вина. В виноделии различают три вида кислотности: фиксированную (FA) или титруемую, летучую (VA) и активную (pH – водородный показатель). Основными фиксированными кислотами, содержащимися в вине, являются винная, яблочная, янтарная, лимонная кислоты. Как правило, фиксированная кислотность пересчитывается на показатель кислотности винной кислоты, так как именно она, наряду с яблочной кислотой, является наиболее значимой в вине. Винная кислота придает вину более мягкие вкусовые ощущения, но в отличие от яблочной кислоты, винная кислота не имеет характерного вкуса. Если сравнить два вина с одинаково высоким уровнем сахара, то образец, имеющий большую кислотность, будет казаться менее сладким и лучше сбалансированным, чем тот, у которого меньший уровень кислотности. Виноделы поддерживают также баланс между содержанием алкоголя и сахара, чтобы избежать приторного чисто сахарного вкуса.

Летучая кислотность (VA) – газообразные кислоты, присутствующие в вине. Это муравьиная, уксусная, пропионовая, масляная, валериановая, каприловая и другие высшие жирные кислоты. Самой важной среди летучих кислот по количеству и значению является уксусная кислота. Все аналитические определения летучей кислотности вин производят в пересчете на уксусную кислоту. Повышение содержания летучих кислот в винах объясняются болезнями вин, а также деятельностью различных болезнетворных бактерий. Самой опасной и в то же время наиболее часто встреча-

Таблица 1. Описательные характеристики вин

Показатель	Красное вино, 1599 образцов				Белое вино, 4898 образцов			
	Среднее значение	Минимальное значение	Максимальное значение	Стандартное отклонение	Среднее значение	Минимальное значение	Максимальное значение	Стандартное отклонение
FA, г/л	8.3	4.6	15.9	1.7	6.9	3.8	14.2	0.8
VA, г/л	0.5	0.1	1.6	0.18	0.3	0.1	1.1	0.1
CA, г/л	0.3	0.0	1.0	0.2	0.3	0.0	1.7	0.1
RS, г/л	2.5	0.9	15.5	1.41	6.4	0.6	31.6	5.1
Ch, г/л	0.09	0.01	0.61	0.05	0.05	0.01	0.35	0.02
FSD, мг/л	15.9	1.0	72.0	10.5	35.3	2.0	289.0	17.0
TSD, мг/л	46.5	6.0	289.0	32.9	138.4	9.0	440.0	42.5
D, г/мл	0.997	0.990	1.004	0.002	0.994	0.987	1.039	0.003
pH	3.3	2.7	4.0	0.15	3.2	2.7	3.8	0.2
SU, г/л	0.7	0.3	2.0	0.2	0.5	0.2	1.1	0.1
Al, об%	10.4	8.4	14.9	1.1	10.5	8.0	14.2	1.2

ющейся болезнью вин является уксусное скисание. При этой болезни этиловый спирт под действием уксусных бактерий окисляется в уксусную кислоту. Слишком много уксусной кислоты в вине может вызвать неприятный вкус уксуса. Изоамиловый спирт и уксусная кислота порождают изоамилацетат, пахнущий персиком, этиловый спирт и муравьиная кислота – этилформиат с запахом малины.

Небольшое количество лимонной кислоты (CA) может усилить свежесть и аромат вина. Остаточный сахар (RS) – количество сахара, оставшегося после брожения. Редко встречаются вина с содержанием сахара менее 1 г/л, а вина с содержанием сахара выше 45 г/л считаются сладкими. Хлориды (Ch) – количество соли NaCl, присутствующей в вине. Порог восприятия различий между опытными образцами в эксперименте определяется внесением солей в дозе 5 мг/л в белые вина и 15 мг/л – в красные. При этом проявляется сильное влияние катиона натрия (Na^+), по крайней мере, в сочетании с Cl^- , на увеличение солёности, длительности вкуса и уменьшение ощущения горечи. В то же время горький вкус усиливается в красных винах при добавлении соли магния, особенно в комби-

нации с хлоридом, что доказывает важную роль этого катиона в восприятии вина дегустаторами и потребителями.

Диоксид серы (сернистый ангидрид, FSD) является основным консервантом, который применяется при изготовлении вин. Его антимикробные, антиоксидантные свойства стабилизируют вино в течение всего времени хранения. Диоксид

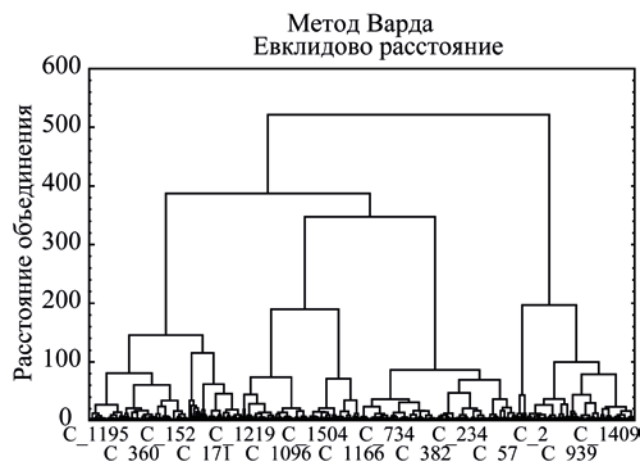


Рис. 2. Дендрограмма результатов агломеративной кластеризации образцов красных вин.

Таблица 2. Распределение образцов вин по кластерам и состав кластеров

Качество вина и состав кластеров					Состав кластеров, %			Качество вина и состав кластеров				Состав кластеров, %		
Кластеры	Количество образцов (1599)	Плохое вино Оценки: 3, 4	Нормальное вино Оценки: 5, 6	Хорошее вино Оценки: 7, 8	Плохое вино Оценки: 3, 4	Нормальное вино Оценки: 5, 6	Хорошее вино Оценки: 7, 8	Количество образцов (4898)	Плохое вино Оценки: 3, 4	Нормальное вино Оценки: 5, 6	Хорошее вино Оценки: 7, 8, 9	Плохое вино Оценки: 3, 4	Нормальное вино Оценки: 5, 6	Хорошее вино Оценки: 7, 8, 9
Красное вино								Белое вино						
Распределение образцов вин по трем кластерам и состав кластеров														
I	509	12	359	138	2.4	70.5	27.1	1799	52	1579	168	3.0	87.7	9.3
II	720	45	609	66	6.3	84.6	9.1	1634	55	1024	552	3.3	62.7	34.0
III	370	6	351	13	1.6	94.9	3.5	1465	76	1052	337	5.2	71.8	23.0
Распределение образцов вин по четырем кластерам и состав кластеров														
I	399	9	299	91	2.2	75.0	22.8	1147	64	883	200	5.6	77.0	17.4
II	314	5	217	92	1.6	69.1	29.3	1167	31	866	270	2.7	74.2	23.1
III	553	45	486	22	8.2	87.8	4.0	1008	43	518	447	4.3	51.4	44.3
IV	333	4	317	12	1.2	95.2	3.6	1576	45	1388	143	2.9	88.1	9.0

серы присутствует в свободной форме (гидрат $\text{SO}_2 \cdot n\text{H}_2\text{O}$) и связанной HSO_3^- , равновесие которых является функцией pH и температуры: $\text{H}_2\text{SO}_3 \leftrightarrow \text{H}^+ + \text{HSO}_3^-$. Соотношение этих форм в вине составляет 10–30%/90–70% (pK_a 1.8).

Общий диоксид серы (TSD) – общая концентрация свободных и связанных форм SO_2 . Повышенная концентрация диоксида серы проявляется в аромате и вкусе вина. Чем больше кислотность вина (меньше pH), тем больше свободного диоксида серы (равновесие сдвигается влево), тем лучше оно защищено.

Плотность (D) вина близка к плотности воды и зависит от концентрации спирта и сахара. pH (активная кислотность) – отрицательный логарифм концентрации свободных ионов водорода $\text{pH} = -\lg[\text{H}^+]$, более строго активности водородных ионов. От величины pH зависит количественное соотношение первичных и вторичных продуктов брожения, склонность вина к окислению, кристаллическим и биологическим помутнениям, подверженность дефектам и сопротивляемость болезням вина. Большинство вин имеют pH от 3 до 4 ($10^{-4} < [\text{H}^+] < 10^{-3}$ моль/л).

Сульфаты (SU) – при окислении диоксид серы переходит в серную кислоту, которая превращается в сульфат калия. Это нежелательный процесс для качества вина. Упрощается вкус, аромат теряет тонкость. Алкоголь (Al об%) – объемное процентное содержание алкоголя в вине. В табл. 1 приведены описательные статистики образцов красных и белых португальских вин верде [24].

Кластерный анализ. Предварительно была проведена z-стандартизация исходных данных:

$$z_i = (x_i - \mu) / \sigma,$$

где z_i – стандартизованный элемент физико-химического показателя; x_i – исходный элемент показателя; μ – среднее арифметическое показателя, σ – стандартное отклонение [25].

На рис. 2 приведена диаграмма объединения в кластеры близких по физико-химическим показателям 1599 образцов красных вин. Аналогичный вид имеет диаграмма агломеративной кластеризации 4898 образцов белых вин. В зависимости от евклидова расстояния (рис. 2) образцы вин объединены в три, четыре и шесть кластеров.

Таблица 3. Распределение образцов вин по шести кластерам и состав кластеров

Качество вина и состав кластеров					Состав кластеров, %			Качество вина и состав кластеров				Состав кластеров, %		
Кластеры	Количество образцов (1599)	Плохое вино Оценки: 3, 4	Нормальное вино Оценки: 5, 6	Хорошее вино Оценки: 7, 8	Плохое вино Оценки: 3, 4	Нормальное вино Оценки: 5, 6	Хорошее вино Оценки: 7, 8	Количество образцов (4898)	Плохое вино Оценки: 3, 4	Нормальное вино Оценки: 5, 6	Хорошее вино Оценки: 7, 8, 9	Плохое вино Оценки: 3, 4	Нормальное вино Оценки: 5, 6	Хорошее вино Оценки: 7, 8, 9
Красное вино								Белое вино						
I	28	1	26	1	3.6	92.9	3.5	979	48	741	190	4.9	75.7	19.4
II	332	8	312	12	2.4	94.0	3.6	1331	31	1163	137	2.3	87.4	10.3
III	302	4	207	91	1.3	68.6	30.1	617	57	532	28	9.2	86.3	4.5
IV	527	43	461	23	8.1	87.5	4.4	919	20	457	442	2.2	49.7	48.1
V	46	1	40	5	2.2	87.0	10.8	948	23	666	259	2.4	70.3	27.3
VI	364	6	273	85	1.6	75.0	23.4	104	4	96	4	3.8	92.3	3.9
Σ	1599	63	1319	217				4898	183	3655	1060			

Таблица 4. Подтверждение распределения плохих, нормальных и хороших вин по кластерам, выделенным методами агломеративной и итеративной кластеризации, %

Число кластеров	Красное вино, 1599 образцов						Белое вино, 4989 образцов					
	MLP	DA	CT	SVM	NBC	kNN	MLP	DA	CT	SVM	NBC	kNN
3	99.94	95.2	95.3	97.3	95.2	91.7	99.88	96.2	97.2	98.0	94.3	93.6
4	99.62	93.1	90.2	97.0	93.4	91.3	99.08	93.9	84.5	96.6	89.4	91.6
6	99.50	93.7	89.9	94.0	84.9	93.7	98.51	92.7	82.1	94.3	86.5	91.9

В табл. 2 и 3 приведены результаты кластеризации образцов вин алгоритмом k -средних (метод Варда) [25]. Объединение поазателей вин в кластеры методом k -средних – эффективный алгоритм [25], имеющий, однако, два существенных недостатка. Во-первых, итоговые результаты чувствительны к начальному случайному выбору центров групп (кластеров). Устранение этого недостатка заключалось в многократном выполнении алгоритма с различным выбором начальных центров кластеров (центроидов). На платформе STATISTICA предлагается три варианта выбора начальных центров кластеров. В работе использован второй вариант – сортировать расстояния и выбрать наблюдения на постоянных интерва-

лах. Второй недостаток – необходимость априори задавать фиксированное число кластеров для разбиения образцов вин, выбор, который, далеко не всегда может быть оптимальным. Поэтому в работе использованы результаты агломеративной кластеризации – разбиение образцов вин на три, четыре и шесть групп (кластеров).

Для подтверждения результатов агломеративной и итеративной кластеризации (табл. 2 и 3) проведены нейросетевой (MLP) и дискриминантный (DA) анализ; применены алгоритмы деревьев классификации (CT), машины опорных векторов (SVM), наивной баесовской классификации (NBC) и ближайших соседей (kNN) [24–26]. Результаты анализов представлены в табл. 4. Лучшую произ-

Таблица 5. Итоги обучения нейросетевых классификаторов

Архитектура	Производительность обучения	Контрольная производительность	Тестовая производительность	Алгоритм обучения	Функция ошибки	Функция активации скрытых нейронов	Функция активации выходных нейронов
Красное вино							
MLP 11-7-3 3 кластера	100	100	99.6	BFGS 44	Энтропия	Тожественная	Софтмакс
MLP 11-13-4 4 кластера	99.9	98.7	99.2	BFGS 30	Энтропия	Тожественная	Софтмакс
MLP 11-14-6 6 кластеров	100	98.3	98.3	BFGS 38	Энтропия	Гиперболическая	Софтмакс
Белое вино							
MLP 11-9-3 3 кластера	100	99.7	99.5	BFGS 92	Энтропия	Логистическая	Софтмакс
MLP 11-5-4 4 кластера	99.3	99.0	98.0	BFGS 123	Энтропия	Гиперболическая	Софтмакс
MLP 11-9-6 6 кластеров	99.0	97.4	97.1	BFGS 115	Энтропия	Гиперболическая	Софтмакс

Таблица 6. Чувствительность обученных нейронных сетей к физико-химическим показателям вин

Сеть	Красное вино										
MLP 11-7-3	TSD	CA	FSD	FA	VA	pH	SU	D	Al	RS	Ch
MLP 11-13-4	TSD	Al	FSD	D	FA	CA	VA	pH	SU	RS	Ch
MLP 11-14-6	FSD	RS	TSD	Al	D	CA	FA	Ch	VA	pH	SU
Сеть	Белое вино										
MLP 11-9-3	FA	pH	RS	FSD	Al	TSD	SU	D	CA	Ch	VA
MLP 11-5-4	FA	pH	Al	FSD	RS	SU	TSD	D	VA	CA	Ch
MLP 11-9-6	Al	pH	Ch	VA	D	FA	SU	TSD	RS	CA	FSD

водительность продемонстрировали нейросетевые модели.

Нейросетевой анализ. В табл. 5 приведены итоги обучения нейросетевых классификаторов MLP (многослойный персептрон) [25, 26]. В табл. 6 приведены последовательности убывания чувствительности (слева направо) обученных нейронных классификаторов к изменению входных физико-химических показателей вин. Полужирным шрифтом выделены показатели, вклад которых в разделяющую мощь нейросетевых моделей определяющий.

Нужно отметить, что анализ чувствительности (табл. 6) позволяет сделать вывод о важности входных переменных для конкретной нейронной сети, но не позволяет надежно определить полезность переменных в процедуре кластеризации, поэтому

выводы о важности переменных носят предварительный характер [25].

Авторами [16] проведен системно-когнитивный анализ зависимости субъективных оценок сомелье качества образцов португальских белых вин от объективных физико-химических показателей. Показано, что качество вина тем выше, чем больше в нем алкоголя (Al), хлорида натрия (Ch), сульфатов (SU) и чем меньше в вине лимонной кислоты (CA), свободного диоксида серы (FSD), общего диоксида серы (TSD), плотность (D) и кислотность (pH).

В табл. 7 приведены диапазоны изменения физико-химических показателей в трех кластерах красных и белых вин для плохих, нормальных и хороших вин. Результаты системно-когнитивного анализа образцов вин [16] и данные табл. 7 являются

Таблица 7. Диапазоны изменения физико-химических показателей образцов красных и белых вин

Показатель	Все образцы вина		Плохое вино, оценки: 3, 4		Нормальное вино, оценки: 5, 6		Хорошее вино, оценки: 7, 8, 9	
	красное вино (1599)	белое вино (4898)	красное вино (63)	белое вино (183)	красное вино (1319)	белое вино (3655)	красное вино (217)	белое вино (1060)
FA	4.5–15.9	3.8–14.2	4.6–12.5	4.2–11.8	4.7–15.9	3.8–14.2	4.9–15.6	3.9–9.2
VA	0.12–1.58	0.08–1.1	0.23–1.58	0.17–1.1	0.16–1.33	0.13–0.97	0.12–0.92	0.08–0.76
CA	0.0–1.0	0.0–1.66	0.0–1.0	0.0–0.88	0.0–0.79	0.0–1.66	0.0–0.76	0.01–0.74
RS	0.9–15.5	0.6–31.6	1.2–12.9	0.7–17.55	0.9–15.5	0.6–31.6	1.2–8.9	0.8–19.25
Ch	0.01–0.61	0.01–0.35	0.05–0.61	0.01–0.29	0.03–0.61	0.01–0.35	0.01–0.36	0.01–0.14
FSD	1–72	2–289	3–41	3–289	1–72	2–131	3–54	5–108
TSD	6–289	9–440	7–119	10–440	6–165	9–344	7–289	34–229
D	0.990–1.004	0.987–1.039	0.993–1.001	0.995–1.001	0.990–1.004	0.987–1.039	0.991–1.003	0.987–1.001
pH	2.74–4.01	2.72–3.82	2.74–3.90	2.83–3.72	2.86–4.01	2.72–3.81	2.88–3.78	2.84–3.82
SU	0.33–2.0	0.22–1.08	0.33–2.0	0.25–0.87	0.37–1.98	0.23–1.03	0.39–1.36	0.22–1.08
AL	8.4–14.9	8.0–14.2	8.4–13.1	8.0–13.5	8.4–14.9	8.0–14.0	9.2–14.0	8.5–14.2

Таблица 8. Возможности нейросетевых алгоритмов (MLP) для подтверждения оценок (%) дегустаторов 5 и 6 (нормальное вино) в трех кластерах

Кластер	Красное вино, 1319 образцов			Белое вино, 3655 образцов		
	MLP	количество образцов	%	MLP	количество образцов	%
I	11-11-2	359 136 (5) 223 (6)	93.3	11-11-2	1579 785 (5) 794 (6)	84.0
II	11-9-2	609 303 (5) 306 (6)	84.4	11-9-2	1024 291 (5) 733 (6)	80.0
III	11-10-2	351 243 (5) 108 (6)	86.0	11-9-2	1052 381 (5) 671 (6)	82.5

ся основой для выделения плохих, нормальных и хороших вин по физико-химическим показателям. Затем для предсказания оценки сомелье задействуются алгоритмы DataMining [25]. В качестве примера в табл. 8 продемонстрирована дискриминирующая мощность нейросетевых алгоритмов (MLP) при разделении образцов нормальных вин с оценкой 5 и оценкой 6 в трех кластерах красных и белых вин.

Вычислительные процедуры следует так же применить к анализу физико-химических показателей образцов плохих, нормальных и хороших

вин, распределенных по четырем (табл. 2) и шести (табл. 3) кластерам, для согласования прогнозной оценки качества вина алгоритмами STATISTICA, использованными в работе.

В заключение следует отметить, что субъективный характер органолептики [5], проблема оптимального выбора физико-химических тестов [3, 7] несбалансированность выборок [23], недостаточная мощность прогнозных моделей усложняют предсказание однозначной оценки сомелье качества образцов вин в 10-балльной шкале по физико-химическим показателям алгоритмами

STATISTICA. Поэтому «несмотря на отдельные субъективные стороны органолептической оценки, ее следует считать основной, а химический, физико-химический и микробиологический методы анализа – вспомогательными» [1].

ИНФОРМАЦИЯ ОБ АВТОРАХ

Бондарев Николай Васильевич, ORCID: <https://orcid.org/0000-0003-2443-1633>

КОНФЛИКТ ИНТЕРЕСОВ

Автор заявляет об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Косюра В.Т., Донченко Л.В., Надыкта В.Д. Основы виноделия: учебное пособие для вузов. М.: Юрайт. 2023. 422 с.
2. Paracelsus T. // 1965. Werke. Bd. 2. Darmstadt. S. 508.
3. Forina M., Armanino C., Casting M., Ubigli M. // Vitis. 1986. Vol. 25. P. 189. doi 10.5073/vitis.1986.25.189-201
4. Sun LX., Danzer K., Thiel G. // Fresenius J. Anal. Chem. 1997. Vol. 359. P. 143. doi 10.1007/s002160050551
5. Ebeler S.E. // Food Rev. Int. 2007. Vol. 17. N 1. P. 45. doi 10.1081/FRI-100000517
6. Legin A., Rudnitskaya V., Lvova L., Vlasov Yu., Di Natale C., D'Amico A. // Anal. Chim. Acta. 2003. Vol. 484. N 1. P. 33. doi 10.1016/S0003-2670(03)00301-5
7. Cortez P., Cerderia A., Almeida A., Matos V., Reis J. // Decision Support Systems. 2009. Vol. 47. N 4. P. 547. doi 10.1016/j.dss.2009.05.016
8. Appalasamy P., Mustapha A., Rizal N., Johari F., Mansor A. // J. Appl. Sci. 2012. Vol. 12. N 6. P. 598. doi 10.3923/jas.2012.598.601
9. Baykal H., Yildirim H.K. // Crit. Rev. Food Sci. Nutr. 2013. Vol. 53. N 5. P. 415. doi 10.1080/10408398.2010.540359
10. Якуба Ю.Ф., Каунова А.А., Темердашев З.А., Титаренко В.О., Халафян А.А. Аналитика и контроль. 2014. Т. 18. № 4. С. 344.
11. Nebot A., Mugica F., Escobet A. // 5th Int. Conf. SIMULTECH. Colmar, France. 2015. P. 501. doi 10.5220/0005551905010507
12. Er Y., Atasoy A. // IJISAE. 2016. Vol. 4 (Special Issue). P. 23. doi 10.18201/ijisae.265954
13. Халафян А.А., Темердашев З.А., Гугучкина Т.И., Якуба Ю.Ф. // Аналитика и контроль. 2017. Т. 21. № 2. С. 161. doi 10.15826/analitika.2017.21.2.010
14. Gupta Y. // Procedia Comput. Sci. 2018. Vol. 125. P. 305. doi 10.1016/j.procs.2017.12.041
15. Ahammed B., Abedin M. // Model Assist. Stat. Appl. 2018. Vol. 13. N 1. P. 85. doi 10.3233/MAS-170420
16. Луценко Е.В., Печурина Е.К., Сергеев А.Э. // Научный журнал КубГАУ. 2019. № 149(05). С. 2. doi 10.21515/1990-4665-149-015
17. Shruthi P. // 1st Int. Conf. ICATIECE. Bangalore, India. 2019. doi 10.1109/ICATIECE45860.2019.9063846
18. Chao Y., Li K., Jia G. // J. Phys. Conf. Ser. 2020. Vol. 1684. N 1. 012067. doi 10.1088/1742-6596/1684/1/012067
19. Kumar S., Agrawal K., Mandan N. // Int. Conf. ICCCI. Coimbatore, India. 2020. doi 10.1109/ICCCI48352.2020.9104095
20. Zhang S., Shao C., Xiao W. // 3rd Int. Conf. ICAIBD. Chengdu, China. 2020. P. 128. doi 10.1109/ICAIBD49809.2020.9137477
21. Мильман Б.Л., Журкович И.К. // ЖАХ. 2020. Т. 75, № 4. С. 316. doi 10.31857/S0044450220020139; Milman B.L., Zhurkovich I.K. // J. Anal. Chem. 2020. Vol. 75. N 4. P. 316. doi 10.31857/S0044450220020139
22. Mor N.S., Asras T., Gal E., Demasia T., Tarab E., Ezekiel N., Nikapros O., Semimufar O., Gladky E., Karpenko M., Sason D., Maslov D., Mor O. // agriRxiv. 2022. P. 1. doi 10.31220/agriRxiv.2022.00125
23. Machine Learning Repository files. <https://archive.ics.uci.edu/ml/datasets/Wine+Quality>
24. Боровиков В.П. Популярное введение в современный анализ данных и машинное обучение на STATISTICA. М.: Горячая линия-Телеком, 2019. 354 с.
25. StatSoft, Inc. Электронный учебник по статистике. М.: StatSoft. WEB: <http://www.statsoft.ru/home/textbook/default.htm>
26. Бондарев Н.В. // ЖОХ. 2021. Т. 91. Вып. 3. С. 449. doi 10.31857/S0044460X21030112; Bondarev N.V. // J. Gen. Chem. 2021. Vol. 91. N 3. P. 409. doi 10.1134/S1070363221030117

Clustering of Organoleptic Quality Evaluations of Red and White Wines by Physicochemical Parameters Using STATISTICA Software

N. V. Bondarev^{a,*}

^a *V.N. Kharkiv National University, Kharkiv, 61022 Ukraine*

^{*}*e-mail: n_bondarev@ukr.net*

Received August 15, 2023; revised August 15, 2023; accepted August 16, 2023

Organoleptic evaluations in a ten-point scale of wine experts and experimental physicochemical parameters of red (1599 samples) and white (4898 samples) wines of Portuguese manufacturers were analyzed using STATISTICA software. Methods of agglomerative and iterative (*k*-means algorithm) clustering revealed the grouping of similar wine samples into three, four and six clusters depending on the Euclidean distance of association. The quantitative filling of clusters with samples of bad wines (grades 3 and 4), normal wines (grades 5 and 6) and good wines (grades 7, 8, 9) was established. Neural network (MLP) and discriminant analyzes (DA) were performed; algorithms of classification trees (CT), support vector machines (SVM), naive Bayesian classification (NBC) and nearest neighbors (kNN) were involved. The best performance was demonstrated by neural network models. Multilayer perceptrons classifiers were trained: for red wines – MLP 11-7-3, MLP 11-13-4, MLP 11-14-6; for white wines – MLP 11-9-3, MLP 11-5-4, MLP 11-9-6. The properties of wines, whose contribution to the separating power of classifiers is decisive, are revealed. The ranges of changes in physicochemical parameters in three clusters of red and white wines for bad, normal and good wines were given.

Keywords: wine, quality, clustering, classification, forecasting, STATISTICA