

Чудинова О.С.,

кандидат экономических наук, доцент

доцент кафедры математических методов и моделей в экономике

Оренбургский государственный университет

Россия, г. Оренбург

Русаков В.Н.,

студент

4 курс, «Финансово-экономический факультет»

Оренбургский государственный университет

Россия, г. Оренбург

АНАЛИЗ ДИФФЕРЕНЦИАЦИИ ДАЧНЫХ (САДОВЫХ) УЧАСТКОВ ОРЕНБУРГСКОЙ ОБЛАСТИ

***Аннотация:** В статье представлены результаты анализа дифференциации дачных (садовых) участков Оренбургской области с использованием методов кластерного анализа, моделей бинарного выбора и деревьев принятия решений. По совокупности показателей, характеризующих коммунальную инфраструктуру и местоположение, осуществлено разбиение дачных (садовых) объединений на два класса. Выработаны рекомендации по использованию математического инструментария решения задач кластеризации и классификации участков.*

***Ключевые слова:** кластеризация, классификация, методы кластерного анализа, деревья принятия решений, логит-модель бинарного выбора, дачные участки.*

***Annotation:** The article presents the results of the analysis of differentiation of suburban (garden) areas of the Orenburg region using the methods of cluster analysis, binary choice models and decision trees. According to the set of indicators*

characterizing the municipal infrastructure and location, the division of country (garden) associations into two classes was carried out. Recommendations on the use of mathematical tools for solving problems of clustering and classification of sites are developed.

***Key words:** clustering, classification, methods of cluster analysis, decision trees, logit model of binary choice, suburban areas.*

В последнее десятилетие существенно возрос интерес к дачным (садовым) участкам. Это обусловлено возможностью получения кредита на приобретение загородной недвижимости и прописки в садовых товариществах. Корректный расчет кадастровой стоимости дачных (садовых) участков играет важную роль в совершенствовании хозяйственного управления земельными ресурсами, позволяет адекватно применять систему налогообложения. Одним из важных этапов оценки стоимости является кластеризация участков, результаты которой позволяют учесть при моделировании стоимости пространственную неоднородность данных [1]. Вопросы кластеризации объектов недвижимости в своих работах рассматривали Безруков В.Б., Господчиков Д.В., Пылаева А.В. [2], Гущина А.А. [3]. Однако в этих работах отсутствует систематизация факторов кластеризации, не затрагиваются вопросы классификации новых участков.

Информационной базой исследования послужила информация о ценах сделок и предложений дачных (садовых) участков Оренбургской области за период с 01.01.2016 г. по 31.03.2018 г. Объем выборки составил 1482 участка. После анализа исходных данных на наличие аномальных значений из выборки были исключены 34 участка.

Учитывая информацию о характеристиках земельных участков, содержащуюся в объявлениях о продаже, мнения экспертов в области оценки недвижимости, а также результаты исследования ведущих специалистов [4]-[6] сформирована система показателей, включающая факторы коммунальной

инфраструктуры и факторы местоположения (таблица 1). Для разбиения дачных (садовых) объединений на однородные группы определены эталонные участки объединений. В качестве эталонных участков дачных (садовых) объединений рекомендуется выбрать существующие участки, удельная рыночная цена которых ближе к среднеарифметическому значению цены в этом объединении. В итоге было отобрано 228 эталонных участков.

Таблица 1.

Система показателей кластеризации дачных (садовых) объединений

Факторы коммунальной инфраструктуры	Факторы местоположения
<p>x₁ - наличие объектов торговли (да/нет);</p> <p>x₂ - наличие возможности круглогодичного проживания (да/нет);</p> <p>x₃ - наличие круглогодичной транспортной доступности (да/нет);</p> <p>x₄ - наличие электроснабжения (да/нет);</p> <p>x₅ - наличие водоснабжения (да/нет);</p> <p>x₆ - наличие газоснабжения (да/нет);</p> <p>x₇ - расположение дачного (садового) объединения относительно рекреационной зоны (да/нет);</p> <p>x₈ - наличие факторов, отрицательно влияющих на стоимость объектов недвижимости (да/нет);</p> <p>x₉ - наличие факторов, положительно влияющих на стоимость объектов недвижимости (да/нет).</p>	<p>x₁₀ - расстояние до ближайшего водного объекта, км;</p> <p>x₁₁ - расстояние до дороги, км;</p> <p>x₁₂ - расстояние до г. Оренбурга, км;</p> <p>x₁₃ - расстояние до центра муниципального образования, км;</p> <p>x₁₄ - расстояние до остановки общественного транспорта, км;</p> <p>x₁₅ - численность населения муниципального образования, человек.</p>

Кластеризация дачных (садовых) объединений проведена методом k-средних. Для измерения расстояний между объектами использовалась метрика Говера, которая применима как к количественным, так и к качественным признакам. Количество классов определено на основе дендрограммы, построенной методом Уорда (рисунок 1).

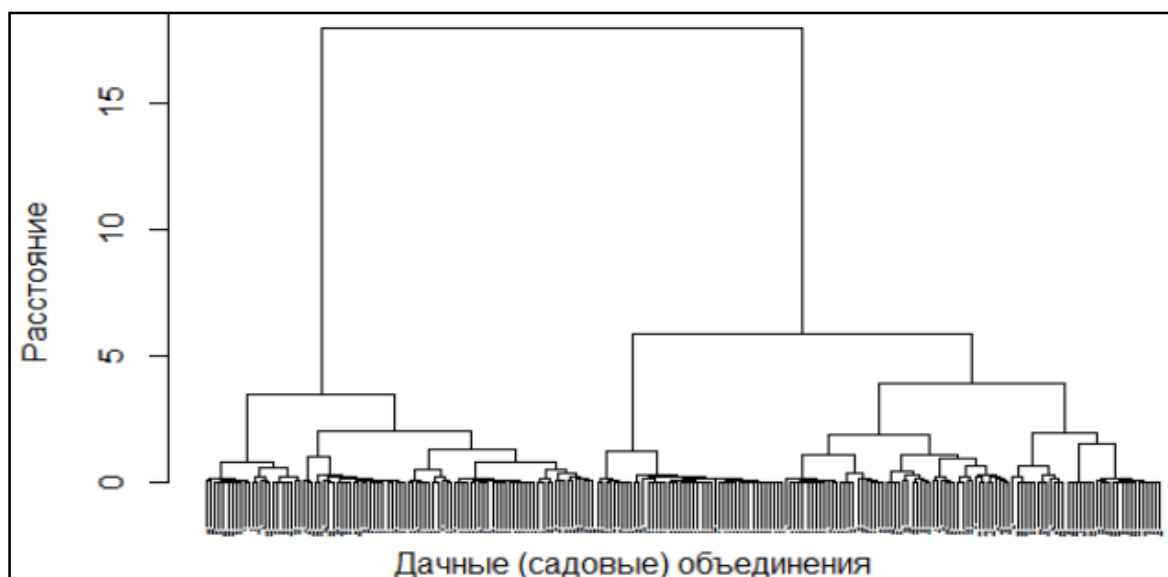


Рисунок 1. Дендрограмма объединения классов методом Уорда

По графику можно сделать вывод о целесообразности разбиения дачных (садовых) объединений на два кластера. Характеристики удельной рыночной цены в каждом классе представлены в таблице 2. Согласно полученным результатам, дачные (садовые) участки, попавшие в первый класс, обладают в среднем наибольшей удельной рыночной ценой по сравнению с участками из второго класса.

Таблица 2.

Количество земельных участков и характеристики удельной цены в классах, построенных методом k-средних

Класс	Количество дачных (садовых) объединений, ед.	Характеристика удельной цены, руб./кв. м.			
		Минимальное значение	Медиана	Среднее значение	Максимальное значение
Первый	183	23,10	218,81	283,01	1123,12
Второй	45	6,92	42,81	85,41	388,15

Средние значения и мода факторов кластеризации представлены в таблице 3.

Таблица 3.

Средние значения (мода) факторов кластеризации в каждом классе

Признак	1 класс	2 класс
x ₁ – наличие объектов торговли (да –1, нет – 0)	0	0
x ₂ – наличие возможности круглогодичного проживания (да –1, нет–0);	1	0
x ₃ – наличие круглогодичной транспортной доступности (да –1, нет–0);	1	0
x ₄ – наличие электроснабжения (да –1, нет – 0);	1	1
x ₅ – наличие водоснабжения (да –1, нет – 0);	0	0
x ₆ – наличие газоснабжения (да –1, нет – 0);	0	0
x ₇ – расположение дачного (садового) объединения относительно рекреационной зоны (да –1, нет – 0);	1	1
x ₈ – наличие факторов, положительно влияющих на стоимость объектов недвижимости (да –1, нет – 0);	0	0
x ₉ – наличие факторов, отрицательно влияющих на стоимость объектов недвижимости (да –1, нет – 0);	0	0
x ₁₀ – расстояние до ближайшего водного объекта, км;	1,214	0,431
x ₁₁ – расстояние до дороги, км;	0,248	0,555
x ₁₂ – расстояние до г. Оренбурга, км;	12,532	218,915
x ₁₃ – расстояние до центра муниципального образования, км;	12,117	7,961
x ₁₄ – расстояние до остановки общественного транспорта, км;	0,595	1,787
x ₁₅ – численность населения муниципального образования, человек	209773	99856,3

Исходя из полученных результатов, первый класс дачных (садовых) объединений можно охарактеризовать как класс с хорошо развитой инфраструктурой, а второй – как класс с плохо развитой инфраструктурой. Для решения задачи классификации дачного (садового) участка, подлежащего

оценке и принадлежащего новому дачному (садовому) объединению, построены две модели: логит-модель бинарного выбора и модель в форме дерева принятия решений. Оценка логит-модели имеет вид:

$$\hat{P}(class = 1 | x) = \frac{\exp(s)}{1 + \exp(s)},$$

$$s = \underset{(1,09)}{3,26} - \underset{(1,28)}{3,24}x_5 + \underset{(1,67)}{4,31}x_7 + \underset{(1,50)}{5,41}x_8 - \underset{(1,42)}{6,73}x_9 - \underset{(1,63)}{5,43}x_{10} + \underset{(0,52)}{1,80}x_{11} -$$

$$\underset{(0,08)}{0,34}x_{13} + \underset{(0,43)}{1,60}x_{14} - \underset{(0,000003)}{0,00001}x_{15},$$

где результирующий показатель class принимает значение 1, если дачный (садовый) участок относится к классу с плохо развитой инфраструктурой и 0, если относится к классу с хорошо развитой инфраструктурой.

Логит-модель бинарного выбора является значимой (наблюдаемое значение статистики Вальда составило Wald chi2(9)=47,84; наблюдаемый уровень значимости $p < 0,05$), псевдо- R^2 для модели составил 0,7652. Построенная логит-модель правильно распознает 95% всех наблюдений, площадь под ROC-кривой составляет 0,98.

В качестве альтернативы логит-модели построены логические правила классификации с использованием деревьев принятия решений [7]. Диаграмма дерева классификации дачных (садовых) объединений, построенная методом CRT в среде RStudio, представлена на рисунке 2. Модель в форме дерева классификации правильно распознает 94,30% всех наблюдений, а площадь под ROC-кривой составляет 0,93.

Построенное дерево классификации имеет 5 терминальных узлов, два из которых прогнозируют принадлежность к первому классу. В построении дерева участвуют только два фактора: численность населения муниципального образования и расстояние до ближайшего водного объекта.

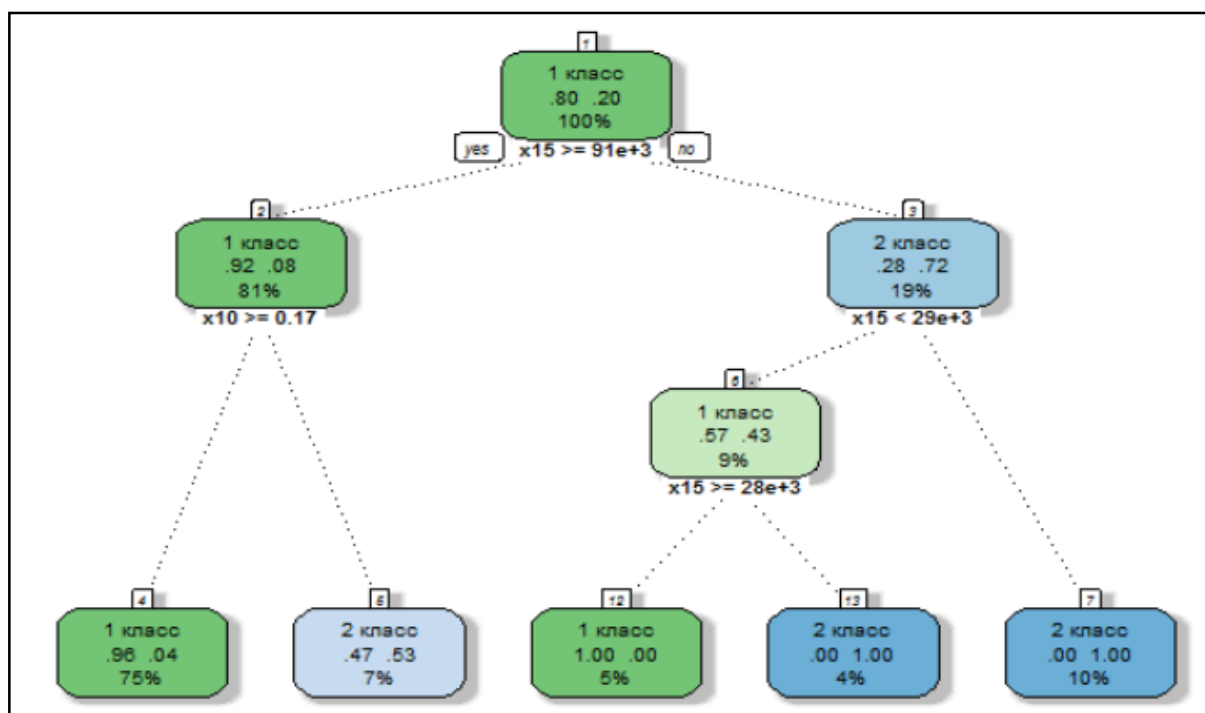


Рисунок 2. Диаграмма дерева классификации дачных (садовых) объединений

Согласно полученной модели дачное (садовое) объединение относится к классу с хорошо развитой инфраструктурой, при выполнении одного из условий:

1) численность населения муниципального образования (x_{15}) больше или равна 91 тыс. человек и расстояние до ближайшего водного объекта (x_{10}) больше или равно 170 м;

2) численность населения муниципального образования (x_{15}) находится на интервале от 28 тыс. до 29 тыс. человек.

Дачное (садовое) объединение относится к классу с плохо развитой инфраструктурой, при выполнении одного из следующих условий:

1) численность населения муниципального образования (x_{15}) меньше 28 тыс. человек;

2) численность населения муниципального образования (x_{15}) находится на интервале от 29 тыс. до 91 тыс. человек;

3) численность населения муниципального образования (x_{15}) больше 91 тыс. человек и расстояние до ближайшего водного объекта меньше 170 м.

Результаты решения задачи классификации дачных (садовых) участков, полученные с помощью метода, основанного на дереве принятия решения, и на основе логит-модели бинарного выбора, согласуются друг с другом и обладают высокой классификационной точностью. Однако в классификации на основе дерева принятия решения участвует всего 2 признака, а в логит-модели бинарного выбора – 9 признаков. Поэтому, с практической точки зрения удобнее применять модель в форме дерева принятия решений.

В связи с низким дискриминирующим свойством третьего правила отнесения участка к классу с плохо развитой инфраструктурой, решение задачи классификации рекомендуется уточнять с помощью логит-модели бинарного выбора. Проведение классификации с использованием двух моделей позволило добиться повышения точности классификации до 99%.

Таким образом, кластеризацию дачных (садовых) объединений рекомендуется проводить методами кластерного анализа с использованием метрики Говера. Решение вопроса о принадлежности нового дачного (садового) объединения к одному из классов рекомендуется осуществлять на основе логических правил классификации, построенных с помощью деревьев принятия решений (методом CRT), уточняя при необходимости классификацию участков из класса с плохо развитой инфраструктурой с помощью логит-модели бинарного выбора.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

1. Русаков, В.Н. Моделирование стоимости дачных (садовых) участков (на примере Оренбургской области) / В.Н. Русаков, О.С. Чудинова // Сборник материалов Международной молодежной научной конференции: «Студенческие научные общества – экономике регионов». III часть – Оренбург: Издательско-полиграфический комплекс ОГУ, 2018. – С. 165-169.
2. Пылаева, А.В. Вопросы кластеризации в процедуре государственной кадастровой оценки земель садоводческих, огороднических и дачных

- объединений вне черты поселений / А.В. Пылаева, В.Б. Безруков, Д.В. Господчиков, // Информационный бюллетень Ассоциации служб кадастра. – 2004. – №4. – С. 44-47.
3. Гущина, А.А. Дифференциация регионального рынка жилья на примере Амурской области / А.А. Гущина // Вестник экономической интеграции. – 2012. – № 10 (55). – С. 95-102.
4. Картунен, О.А. Факторы, влияющие на формирование стоимости // Журнал правовых и экономических исследований. – 2014. – №4. – С.133-134.
5. Ротова, Т.В. Специфика рынка земли и факторы, определяющие цену на землю / Т.В. Ротова // Вестник Кубанского государственного университета. – 2012. – № 5. – С. 7-21.
6. Баева, И.А. Формирование кадастровой оценки земель садоводческих, огороднических и дачных товариществ в Тамбовской области / И. А. Баева и др. // Актуальные проблемы рационального использования земельных ресурсов: материалы Всерос. науч.-практ. конф. (17 мая 2017г.). –Лесниково: Кург.ГСХА, 2017. – С.81-84.
7. Груздев, А.В. Прогнозное моделирование в IBM SPSS Statistics, R и Python: метод деревьев решений и случайный лес. – М.:ДМК Пресс, 2018. – 642 с.