

Назарова Анна Сергеевна, магистрант кафедры информационных и сетевых технологий, Ярославский государственный университет им. П. Г. Демидова, РФ, г. Ярославль

**ПРИМЕНЕНИЕ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ ИЗУЧЕНИЯ
РАСПРОСТРАНЕНИЯ БОРЩЕВИКА
СОСНОВСКОГО НА ТЕРРИТОРИИ ЯРОСЛАВСКОЙ ОБЛАСТИ**

Аннотация

В статье рассматривается проблема распространения борщевика Сосновского, который представляет угрозу для сельского хозяйства. Целью исследования является применение методов кластерного анализа на примере Ярославской области для разделения групп муниципальных образований, различающихся по различным метрикам и степени заражения. Были использованы данные региональной информационной аналитической системы (РИАС) за 2025 год. Методом k-средних с предварительной обработкой данных было выделено три кластера: «зона внимания», «зона повышенного внимания» и «зона больших площадей». Полученная типология может служить основой для планирования мероприятий по борьбе с борщевиком и создания региональной системы мониторинга.

Annotation

The article examines the problem of Sosnowsky's hogweed distribution, which poses a threat to agriculture. The aim of the study is to apply cluster analysis methods using the example of Yaroslavl Region to divide groups of municipal entities that differ by various metrics and degree of infestation. Data from the Regional Information and Analytical System (RIAS) for 2025 were used. Using the k-means method with preliminary data preprocessing, three clusters were identified: «attention zone», «increased attention zone», and «large area zone». The resulting

typology can serve as a basis for planning measures to combat hogweed and creating a regional monitoring system.

Ключевые слова: борщевик Сосновского, кластерный анализ, k-средние, Ярославская область, Python.

Keywords: Sosnowsky's hogweed, cluster analysis, k-means, Yaroslavl Region, Python

Борщевик Сосновского является одним из опасных инвазивных видов растений на территории России. Его быстрое распространение превращается в серьёзную проблему. Он наносит значительный ущерб сельскому хозяйству и представляет угрозу для здоровья населения. Особенно остро проблема ощущается в регионах Центрального федерального округа.

Ярославская область, являясь важным туристическим центром Золотого кольца России, входит в число наиболее заражённых регионов. По данным Геопортала Ярославской области на 2025 год, было выявлено и зафиксировано более 6200 гектаров заражённых территорий, находящихся в частной и муниципальной собственности. В связи с этим возрастает необходимость системного анализа данных и разработки эффективных мер борьбы.

Целью данного исследования является применение методов кластерного анализа для выявления групп муниципальных образований Ярославской области, различающихся по площади заражения, числу проектов по ликвидации борщевика и доле поражённых земель для последующего исследования.

Исходные данные получены из Региональной информационной аналитической системы (РИАС). Таблица содержит сведения об объектах (заявках на устранение борщевика), муниципальном образовании, этапе работ, площади заражения и других характеристиках. Общее число записей – 5187. (рис. 1)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Муниципальное образование	Код объекта	Название	Поселение ГО, СП, ПП	Площадь	Этап	Приоритет	%	Местность	Ведомство	Форма собственности	
2	0	Гаврилов-Ямский МО	3148279	Гаврилов-Ямский	Митинское СП	3600	Запланирован	Высокий		0 д. Ульяново-	ОМСУ	неразграниченная	
3	1	Гаврилов-Ямский МО	3148312	Гаврилов-Ямский	Митинское СП	15400	Запланирован	Высокий		0 с. Осенево о	ОМСУ	муниципальная	
4	2	Гаврилов-Ямский МО	3148309	Гаврилов-Ямский	Митинское СП	100800	Запланирован	Высокий		0 д. Артемиха	ОМСУ	муниципальная	
5	3	Гаврилов-Ямский МО	3148305	Гаврилов-Ямский	Митинское СП	77000	Запланирован	Высокий		0 д. Чайкино п	ОМСУ	муниципальная	
6	4	Гаврилов-Ямский МО	3148293	Гаврилов-Ямский	Митинское СП	79700	Запланирован	Высокий		0 д. Вакуриха	ОМСУ	неразграниченная	
7	5	Гаврилов-Ямский МО	3148292	Гаврилов-Ямский	Митинское СП	49100	Запланирован	Высокий		0 д. Вакуриха	ОМСУ	муниципальная	
8	6	Гаврилов-Ямский МО	3148290	Гаврилов-Ямский	Митинское СП	97	Запланирован	Высокий		0 д. Матвейка	ОМСУ	муниципальная	
9	7	Гаврилов-Ямский МО	3148289	Гаврилов-Ямский	Митинское СП	6700	Запланирован	Высокий		0 д. Кадищи	ОМСУ	муниципальная	
10	8	Гаврилов-Ямский МО	3148288	Гаврилов-Ямский	Митинское СП	11200	Запланирован	Высокий		0 д. Кадищи	ОМСУ	муниципальная	
11	9	Гаврилов-Ямский МО	3148287	Гаврилов-Ямский	Митинское СП	600	Запланирован	Высокий		0 д. Слобода	ОМСУ	муниципальная	
12	10	Гаврилов-Ямский МО	3148285	Гаврилов-Ямский	Митинское СП	980	Запланирован	Высокий		0 д. Слобода	ОМСУ	неразграниченная	
13	11	Гаврилов-Ямский МО	3148283	Гаврилов-Ямский	Митинское СП	8300	Запланирован	Высокий		0 д. Слобода	ОМСУ	неразграниченная	
14	12	Гаврилов-Ямский МО	3148282	Гаврилов-Ямский	Митинское СП	5000	Запланирован	Высокий		0 д. Слобода	ОМСУ	муниципальная	
15	13	Гаврилов-Ямский МО	3148280	Гаврилов-Ямский	Митинское СП	9000	Запланирован	Высокий		0 д. Ульяново-	ОМСУ	неразграниченная	
16	14	Гаврилов-Ямский МО	3148313	Гаврилов-Ямский	Митинское СП	2600	Запланирован	Высокий		0 д. Ульяново	ОМСУ	муниципальная	
17	15	Гаврилов-Ямский МО	3148278	Гаврилов-Ямский	Митинское СП	780	Запланирован	Высокий		0 с. Пружинин	ОМСУ	муниципальная	
18	16	Ярославский МО	3147718	участок в ДНП Нов	Заволжское СП	460	Запланирован	Высокий		0 земельный у	ОМСУ	частная	
19	17	Ярославский МО	3147717	участок в ДНП Нов	Заволжское СП	370	Запланирован	Высокий		0 земельный у	ОМСУ	частная	
20	18	Ярославский МО	3147716	участок в ДНП Нов	Заволжское СП	290	Запланирован	Высокий		0 земельный у	ОМСУ	частная	
21	19	Ярославский МО	3147715	участок в ДНП Нов	Заволжское СП	530	Запланирован	Высокий		0 земельный у	ОМСУ	частная	
22	20	Ярославский МО	3147714	участок в ДНП Нов	Заволжское СП	490	Запланирован	Высокий		0 земельный у	ОМСУ	частная	
23	21	Ярославский МО	3147713	участок в ДНП Нов	Заволжское СП	480	Запланирован	Высокий		0 земельный у	ОМСУ	частная	
24	22	Ярославский МО	3147712	участок в ДНП Нов	Заволжское СП	470	Запланирован	Высокий		0 земельный у	ОМСУ	частная	
25	23	Ярославский МО	3147711	участок в ДНП Нов	Заволжское СП	570	Запланирован	Высокий		0 земельный у	ОМСУ	частная	
26	24	Ростовский МО	3145719	р.п. Петровское, ул	СП Петровское	2351	Запланирован	Высокий		0 р.п. Петровск	ОМСУ	муниципальная	
27	25	Ярославский МО	3145455	Григорьевское,кар	Заволжское СП	0	Запланирован	Высокий		0 Григорьевск	ОМСУ	частная	
28	26	Ярославский МО	3145454	Григорьевское, от	Заволжское СП	7600	Запланирован	Высокий		0 Григорьевск	ОМСУ	муниципальная	
29	27	Ярославский МО	3145453	д. Григорьевское, (	Заволжское СП	960	Запланирован	Высокий		0 д. Григорьев	ОМСУ	муниципальная	
30	28	Гаврилов-Ямский МО	3148342	Гаврилов-Ямский	Митинское СП	3600	Запланирован	Высокий		0 д. Вакуриха	ОМСУ	муниципальная	

Рисунок 1

Для анализа данные были сгруппированы по муниципальным образованиям (МО) (рис. 2). Для каждого МО были рассчитаны:

- количество проектов (объектов);
- суммарная площадь заражения;
- доля заражённых земель по отношению к общей площади МО (%).

МО/ГО	Площадь МО	Количество объектов	Площадь поражения, кв. м	% поражения
Большесельский МО	13331100	<u>79</u>	909180	7
Борисоглебский МО	17480300	<u>32</u>	1975597.1	11
Брейтовский МО	21545500	<u>213</u>	574032.81	3
Гаврилов-Ямский МО	11199600	<u>223</u>	2702713.74	24
Даниловский МО	22115700	<u>221</u>	2363440.7	11
Любимский МО	19666300	<u>288</u>	2284742.92	12
Мышкинский МО	11112100	<u>157</u>	1688762.47	15
Не указан	0	<u>751</u>	61180.61	0
Некоузский МО	19544500	<u>353</u>	5888550.53	30
Некрасовский МО	13636700	<u>321</u>	5414318.62	40
Первомайский МО	22269000	<u>117</u>	5179408.31	23
Переславль-Залесский МО	31306700	<u>274</u>	6680987	21
Пошехонский МО	43831100	<u>96</u>	2531303	6
Ростовский МО	20818200	<u>281</u>	1477175.27	7
Рыбинский МО	31416900	<u>107</u>	254653.87	1
Тутаевский МО	14513400	<u>1</u>	150	0
Угличский МО	25683300	<u>380</u>	4197552	16
Ярославский МО	19235700	<u>1007</u>	16007168.51	83
г. Рыбинск	1014200	<u>126</u>	489142	48
г. Ярославль	2058000	<u>161</u>	1383620	67
Итого	361778300	5188	62063679.45	17

Рисунок 2

В ходе первичной обработки данных выявлены следующие недостатки:

- присутствие МО без названия (вероятно, техническая ошибка внесения);
- аномально высокая доля заражения (83%) для Ярославского МО, что связано, скорее всего, с неправильным указанием площади в каком-то из проектов.

Данные недостатки свидетельствуют об отсутствии строгого регламента заполнения данных. Поэтому на следующем этапе подготовки данных для кластерного анализа были исключены:

- записи с отсутствующим названием МО;
- проекты с площадью заражения более 500 000 га (выбросы, явно ошибочные).

После данных преобразований и фильтрации по проценту заражения была получена таблица, пригодная для кластеризации:

Таблица 1.

Для разбиения муниципальных образований на кластеры была написана программа на языке Python с использованием методом k-средних, реализованный в библиотеке Scikit-Learn. Кластеризация проводилась по следующим признакам:

МО	Всего объектов	Площадь поражения	Площадь МО	Процент поражения
Г. Ярославль	161	1383620.00	2058000	67.23
Ярославский МО	1002	10450911.60	19235700	54.33
Г. Рыбинск	126	489142.00	1014200	48.23
Некрасовский МО	319	5414288.62	13636700	39.70
Некоузский МО	353	5888550.53	19544500	30.13
Переславль-Залесский МО	274	6680987.00	31306700	21.34
Первомайский МО	116	4668408.31	22269000	20.96
Гаврилов-Ямский МО	222	2053615.74	11199600	18.34
Угличский МО	380	4197552.00	25683300	16.34
Мышкинский МО	157	1688762.47	11112100	15.20
Любимский МО	288	2284742.92	19666300	11.62
Даниловский МО	221	2363440.70	22115700	10.69
Борисоглебский МО	31	1436583.43	17480300	8.22
Ростовский МО	281	1477175.27	20818200	7.10
Большесельский МО	79	909180.00	13331100	6.82
Пошехонский МО	96	2531303.00	43831100	5.78
Брейтовский МО	213	574032.81	21545500	2.66
Рыбинский МО	107	254653.87	31416900	0.81

МО	Всего объектов	Площадь поражения	Площадь МО	Процент поражения
Тутаевский МО	1	150.00	14513400	0.00

- количество объектов в МО;
- средняя площадь заражения по округу;
- доля заражённой территории (%).

Так как суммарная площадь поражения варьируется от маленьких значений до больших, было выполнено логарифмирование этого показателя для уменьшения влияния экстремально больших значений.

Число кластеров выбрано равным трём, так как два кластера дают слишком грубое деление, а четыре – уже избыточно.

Таблица 2.

МО	Всего объектов	Площадь поражения	Площадь МО	Процент поражения	Кластер
Большесельский МО	79	909180.00	13331100	6.82	0
Брейтовский МО	213	574032.81	21545500	2.66	0
Гаврилов-Ямский МО	222	2053615.74	11199600	18.34	0
Даниловский МО	221	2363440.70	22115700	10.69	0
Мышкинский МО	157	1688762.47	11112100	15.20	0
Любимский МО	288	2284742.92	19666300	11.62	0
Тутаевский МО	1	150.00	14513400	0.00	0
Угличский МО	380	4197552.00	25683300	16.34	0
Рыбинский МО	107	254653.87	31416900	0.81	0
Ростовский МО	281	1477175.27	20818200	7.10	0
Некрасовский МО	319	5414288.62	13636700	39.70	1
Некоузский МО	353	5888550.53	19544500	30.13	1
Г. Рыбинск	126	489142.00	1014200	48.23	1
Г. Ярославль	161	1383620.00	2058000	67.23	1
Ярославский МО	1002	10450911.60	19235700	54.33	1
Борисоглебский МО	31	1436583.43	17480300	8.22	2
Переславль-Залесский МО	274	6680987.00	31306700	21.34	2

МО	Всего объектов	Площадь поражения	Площадь МО	Процент поражения	Кластер
Первомайский МО	116	4668408.31	22269000	20.96	2
Пошехонский МО	96	2531303.00	43831100	5.78	2

Средние показатели по кластерам получились такими:

Таблица 3.

Кластер	Всего объектов	Площадь поражения	Процент заражения
0	194.90	7167.10	8.96
1	392.20	11312.04	47.92
2	129.25	34334.30	14.07

Результаты кластеризации были интерпретированы:

Таблица 4.

Для визуализации многомерных данных использован метод главных компонент, который позволяет представить многомерные данные в двумерном пространстве (рис. 3).

Кластер	Характеристика	Интерпретация
Кластер 1	Относительно небольшая доля заражения и небольшое количество объектов	Зона внимания – требуется регулярный мониторинг ситуации
Кластер 2	Высокая доля заражения и большое количество объектов	Зона повышенного внимания – требуются срочные меры для борьбы с борщевиком
Кластер 3	Маленькое количество объектов, но большая площадь заражения	Зона крупных очагов – требуются проекты на больших территориях

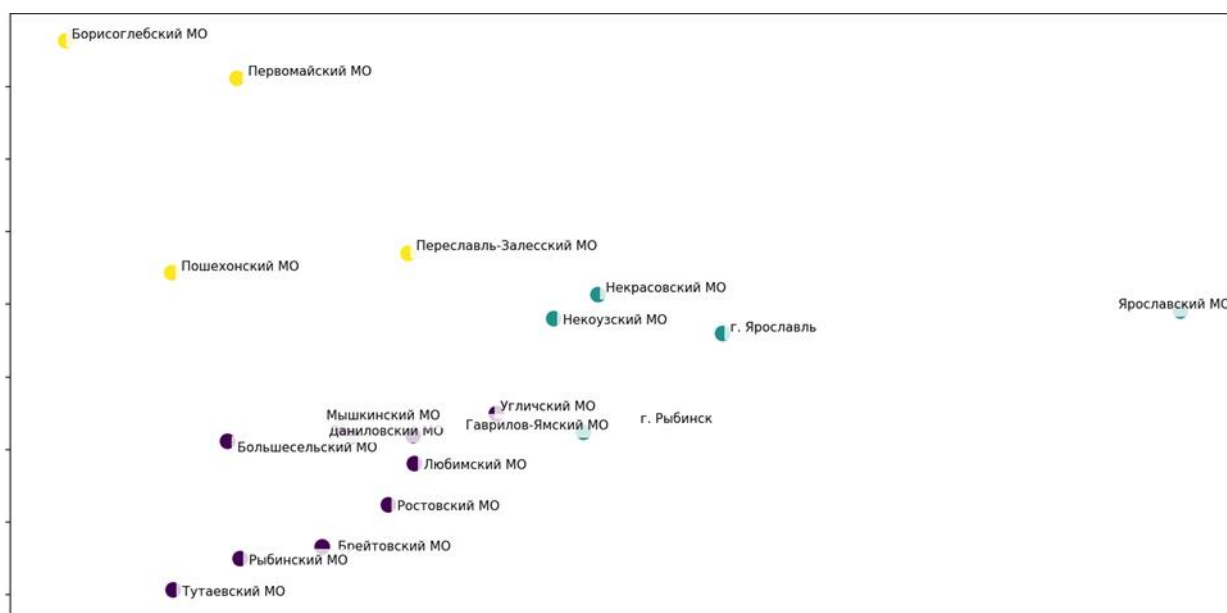


Рисунок 3

Полученные группы позволяют выделить муниципальные образования, которые требуют различных подходов к планированию мероприятий по уничтожению борщевика. Например, для зоны внимания достаточно локальных обработок и мониторинга. Для зоны повышенного внимания требуются системных решения: ежегодные обработки и финансирование из регионального бюджета. Для зоны крупных очагов предлагается разработка программы с участием собственников земель.

Выявленные на этапе предобработки ошибки и неточности в данных (отсутствие названий МО, аномальные значения площадей) указывают на необходимость введения чёткого регламента заполнения электронных заявок. Это повысит точность последующего анализа.

Для мониторинга за распространением борщевика разработан интерактивный дашборд с использованием React.js и библиотеки Vite. Он позволяет визуализировать данные и отслеживать динамику. (рис. 4)

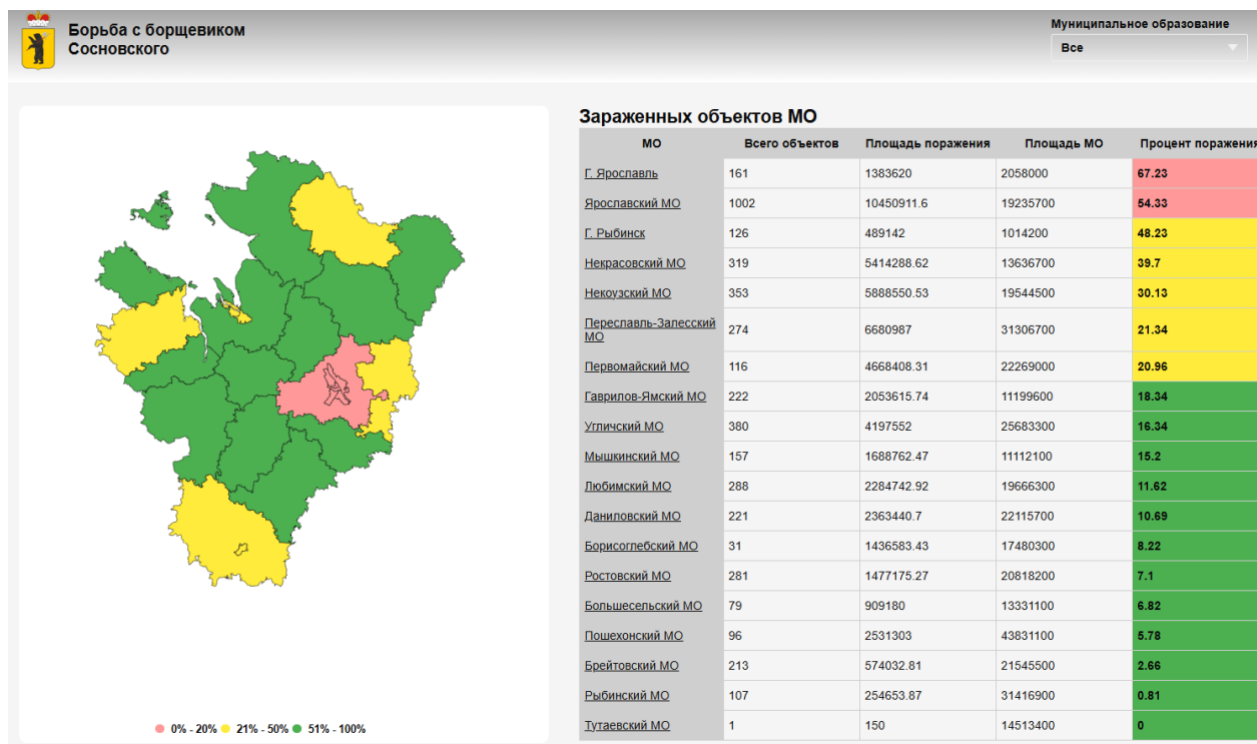


Рисунок 4

Данное исследование показало, что кластерный анализ является эффективным инструментом для выявления неоднородности в распространении борщевика Сосновского по муниципальным образованиям Ярославской области.

В дальнейшем планируется разработка региональной системы для подачи заявок на устранение борщевика Сосновского от граждан. В её состав будут входить: бот в мессенджере МАХ, предназначенный для приёма и обработки заявлений от населения; административная панель для управления заявлениями (изменения статусов обработки заявления, назначение исполнителей); интерактивный дашборд, предоставляющий гражданам актуальную информацию о ходе выполнения работ и текущей ситуации по региону.

Литература

1. Галлатин К., Элбон К. Машинное обучение с использованием Python. Сборник рецептов. – Астана: АЛИСТ, 2024. – 448 с.
2. Полетаев А.Ю., Спиридонова Е.М. Оценка эффективности метода оптимизации инвестиционного портфеля с использованием иерархической кластеризации // Заметки по информатике и математике: сб. науч. ст. – 2020. – С. 247–253.
3. Садовникова Т.П., Ульянкина Т.Д., Снакин В.В. Опасный интродуцент: борщевик Сосновского // Использование и охрана природных ресурсов в России. – 2018. – № 3. – С. 63–67.
4. Стефанов С. React.js. Быстрый старт. – СПб.: Питер, 2017. – 304 с.
5. Scikit-learn: K-Means Clustering [Электронный ресурс]. // Scikit-learn Documentation. – URL: <https://scikit-learn.org/stable/modules/clustering.html#k-means> (дата обращения: 12.05.2026).
6. Vite [Электронный ресурс]: официальная документация. – URL: <https://vite.dev/guide/> (дата обращения: 16.05.2026).