

УДК

---

**КЛАСТЕРНЫЙ АНАЛИЗ ИННОВАЦИОННОЙ АКТИВНОСТИ  
РЕГИОНОВ РФ НА ОСНОВЕ МЕТОДА MEAN SHIFT**

---

**Московкин Владимир Михайлович**

доктор географических наук  
заместитель директора Центра развития публикационной активности и  
научно-издательской деятельности  
профессор кафедры мировой экономики  
Белгородского государственного национального исследовательского университета  
г. Белгород, Российская Федерация  
moskovkin@bsu.edu.ru

**Сизьунго Муненге**

научный стажер кафедры экономики и  
моделирования производственных процессов НИУ «БелГУ»  
Белгород

**Голиков Николай Александрович**

независимый исследователь  
фрилансер, Minitab Inc., США

**Аннотация**

---

Прделана кластеризация регионов России за десятилетний период времени на основе метода Mean Shift, основанного на объектно-ориентированном программировании и алгоритме самообучения, который в отличие от популярного метода K-Means, не требует предварительного задания количества кластеров. В качестве исходной базы данных для кластеризации использованы 16 инновационных и 2 макроэкономических индикатора Росстата. В результате кластеризации, поддержанной картографированием, идентифицированы четыре группы регионов: "среднячки", "отстающие", "лидеры", "аномально эффективные". Выделены также шесть поведенческих типов регионов.

---

**Ключевые слова:** кластеризация регионов, инновационная активность, регионы России, картосхема, Mean Shift, K-Means.

---

**CLUSTER ANALYSIS OF INNOVATION ACTIVITY OF THE RUSSIAN  
FEDERATION REGIONS ON THE BASIS OF MEAN SHIFT METHOD****Vladimir M. Moskovkin**

Doctor of Geographical Sciences  
Deputy Director for Development and scientific Publication Activity Center  
Professor of the World Economy Department  
Belgorod State National Research University

Belgorod, Russian Federation  
moskovkin@bsu.edu.ru

### **Sizyoongo Munenge**

Research fellow

Department of Economics and Modeling of Production Processes

Belgorod State University

Belgorod

### **Nikolay A. Golikov**

Independent researcher,

Freelancer, Minitab Inc., USA

---

#### **ABSTRACT**

---

The article presents clustering of Russian regions for a ten-year period on the basis of the Mean Shift method based on object-oriented programming and self-learning algorithm, which, unlike the popular K-Means method, does not require a preliminary specification number of clusters. 16 innovation indicators and 2 macroeconomic indicators of Rosstat were used as the initial database for clustering. As a result of the clustering supported by mapping, four groups of regions were identified: "average", "lagging", "leaders", and "anomalously effective". The paper also highlights six behavioral types of regions.

---

**Key words:** regional clustering, innovation activity, Russian regions, map chart, Mean Shift, K-Means.

---

#### **Введение**

В нашей предыдущей статье мы рассмотрели три метода картографической кластеризации инновационной активности регионов России, отличающиеся способом построения шкал [1]. При такой кластеризации мы использовали один единственный показатель, который являлся интегральным показателем инновационной активности. В нем уже были агрегированы специальным образом все частные инновационные индикаторы. В тоже время при кластеризации различных объектов обычно используют не такие одномерные (скалярные) интегральные показатели объектов, а исходные многомерные объекты, описываемые их индикаторами. На таких многомерных объектах, представляемых векторами, обычно проводится их кластеризация. Суть такой кластеризации состоит в выделении групп объектов взаимные расстояния, между которыми минимальны, а расстояния между центрами различных групп максимальны. В качестве расстояний между объектами в группах и центрами групп чаще всего используются евклидовы расстояния. Наиболее часто экономисты для такой кластеризации используют иерархические методы типа методов польской таксономии или метод K-Means, относящийся к неиерархическим методам. В нашем исследовании мы использовали метод Mean Shift, который относится к методам машинного обучения и не требует заранее, в отличие от метода K-Means, задания количества кластеров.

#### **Методология исследований. Результаты и обсуждение**

В качестве системы инновационных индикаторов брались те же индикаторы, что использовались в работе [1]. Приведем их ниже (табл.1).

Таблица 1.

Система инновационных индикаторов инновационного развития

№	Название индикатора
1	Количество организаций, выполняющих научные исследования и разработки
2	Численность персонала без ученых степеней, занятых НИОКР
3	Численность исследователей, имеющих ученую степень кандидата наук, по субъектам РФ
4	Численность исследователей, имеющих ученую степень доктора наук, по субъектам РФ
5	Внутренние затраты на научные исследования и разработки (млн. руб.)
6	Численность аспирантов по субъектам Российской Федерации
7	Численность докторантов по субъектам Российской Федерации
8	Количество патентов, выданных на изобретения
9	Количество патентов, выданных на полезные модели
10	Количество разработанных передовых производственных технологий по субъектам Российской Федерации
11	Количество используемых передовых производственных технологий по субъектам Российской Федерации
12	Удельный вес организаций, использовавших интернет по субъектам Российской Федерации (в процентах % от общего числа обследованных организаций соответствующего субъекта РФ)
13	Удельный вес организаций, осуществлявших технологические, организационные, маркетинговые инновации в отчетном году, в общем числе обследованных организаций (%)
14	Затраты на технологические инновации организаций (млн. руб.)
15	Объем инновационных товаров, работ, услуг (отгружено товаров собственного производства, выполнено работ и услуг собственными силами, в том числе инновационных товаров, работ, услуг) (млн. руб.)
16	Удельный вес инновационных товаров, работ, услуг в общем объеме отгруженных товаров, выполненных работ, услуг по субъектам РФ (%)

То есть исходной базой данных для кластеризации служил старый массив данных (значения 16 индикаторов за 10 лет по 83 регионам РФ), к ним только добавлены два макроэкономических индикатора – численность населения на конец года в тысячах человек и валовый региональный продукт в текущих ценах (млн. руб.).

Загружаем необходимые для работы библиотеки (Pandas, Nump ScikitLearn, Matplotlib стандартный набор библиотек применяемых в Data Science) языка Python, а также включаем директиву `%matplotlib inline` необходимую для отображения изображений построенных посредством библиотеки Matplotlib.

Читаем источник данных (файл в формате .xlsx) с помощью метода `read_excel` библиотеки Pandas предназначенной для работы с данными, прежде всего для операций с DataFrames основным способом представления данных используемым различным программным обеспечением в области Machine Learning / Data Science. Каждый лист Excel-я читаем отдельно получая для каждого листа отдельный датафрейм.

Для дальнейших операций объединяем полученные датафреймы в список, который в свою очередь состоит из подсписков содержащих собственно сам датафрейм и ассоциированные с ним названия максимально соответствующее русским названиям экселевского листа. В дальнейшем, эти названия станут именами столбцов датафреймов содержащих информацию разбитую по годам.

Отдельно сохраняем соответствующие русскоязычные названия в виде списка, в том же порядке в котором идут датафреймы в их списке:

```
features_names_ru = [
```

- количество организаций, выполняющие научные исследования и разработки,  $n$ ;

...

- численность персонала без ученых степеней, занятых НИОКР, человек;

-внутренние затраты на научные исследования и разработки \n' миллионов рублей'. ]

Корректируем содержащие ошибочные символы (некорректно внесенные в исходные данные) в названиях некоторых административных единиц. Так же некоторые названия заменяем на другие варианты названий, которые встречаются в иных источниках данных, используемых в данной работе (списке кодов регионов в географических данных).

Создаём список «годовых» датафреймов, каждый из которых содержит полный список показателей (features в терминах Data Science) для каждого региона (столбцы показатели, строки регионы). Используем стандартную операцию Pandas-овских датафреймов merge (для объединения «показательных» датафреймов в «годовые»). В качестве столбца используемого для объединения используем столбец "Region". То есть объединяем все показатели по данному региону за год.

Вспомогательные функции для корректировки некорректно представленных значений (например, содержащих в качестве десятичного разделителя запятую) в датафреймах импортируем в стандартный класс библиотеки ScikitLearn для приведения значений к одному и тому же масштабу. Создаем экземпляр этого класса. Корректируем вручную одно из неправильных значений, а ко всему остальному применяем функции корректировки.

Объединяем список «годовых» датафреймов в один датафрейм с иерархическим индексом «Год/Регион». В дальнейшем будем производить кластеризацию, используя именно этот датафрейм. Кластер в общем случае объединяет «в чем то схожие» регионы, причем кластер при этом имеет временную (ударение на «у») протяжённость. То есть каждый кластер простирается от 2005 до 2014-го года, и один и тот же регион в разные годы может принадлежать к различным кластерам (например, в 2005 г. к «первому», в 2006 г. ко «второму», а в остальные годы к «четвертому»), а может и все время оставаться в одном и том же кластере.

Приводим к единому масштабу все данные, используя ранее созданный экземпляр класса MinMaxScaler.

Термин «класс» используется в двух различных значениях:

1. Концепция ООП программирования, широко используемая в большинстве современных языков программирования, служащая для объединения программного кода в некие семантически/логически/функционально связанные сущности, и предоставляющая средства для оперирования с этими сущностями.

2. Как класс в задачах классификации машинного обучения. В данном случае смысл термина класс близок к бытовому его восприятию (например классом в таком понимании является сорт цветка - роза, тюльпан, и т.д. или тип транспортного средства - автомобиль, мотоцикл, самолет и т.д.).

Алгоритмы машинного обучения в библиотеке ScikitLearn оформлены в виде классов, получающих на входе (при создании экземпляра класса) параметры определяющие работу соответствующего алгоритма. Далее при вызове метода fit созданный экземпляр класса получает значения используемые для обучения в виде массивов библиотеки NumPy. Извлекаем из дата-фрейма соответствующий массив NumPy, просто обращаясь к полю датафрейма values.

В качестве алгоритма кластеризации используем MeanShift [2] [3].

Для неформального понимания того как он работает рассмотрим самый простой случай. Пусть у нас есть некий набор точек в одномерном пространстве, то есть попросту говоря на прямой. Набор точек произвольный, единственное ограничение налагаемое на

него конечное количество точек. Понятно, что в общем случае в некоторых местах точки могут быть расположены чаще (более плотно), в других - реже (менее плотно). Таких уплотнений в общем случае может быть несколько, и интуитивно понятно, что для каждого такого уплотнения можно найти центр, и вычислить его координаты (а координат у нас в рассматриваемом случае одна). Понятно, что в частном случае равномерно расположенных точек в качестве центра можно взять просто среднее арифметическое значение координат точек (расстояний от 0). Таким же образом поступаем и в многомерном случае.

Далее мы принимаем следующее утверждение: центры «уплотнений» и есть центры кластеров, а точки входящие в данное «уплотнение» относятся к одному кластеру. Интуитивно понятно, что если точки где-то расположены более тесно, то их можно считать в чем то близкими. Здесь можно сделать первое формальное уточнение. Что означают точки «относящиеся к данному уплотнению»? Очевидно, что это те точки, расстояние от которых до центра данного уплотнения меньше, чем до центра любого другого уплотнения.

Теперь нам надо как то превратить неформальный термин «уплотнение» в математический вид. Во первых, раз мы хотим как-то измерить эту самую «степень уплотнённости», то мы должны поставить в соответствие этой самой «степени уплотнённости» некое число. И если где-то точки расположены более плотно, то там это число должно быть больше, чем там, где точки расположены менее плотно. А что такое «степень уплотнённости»? Собственно само слово «плотность» уже содержит в себе ответ. Это отношение массы чего-то заключённой в каком-то объёме к этому объёму. Ну, или в более обобщённом виде (и более строго), это отношение количества чего-то заключённого в каком-то замкнутой области  $n$ -мерного пространства к  $n$ -мерному объёму этой области. В случае опять же одного измерения  $n$ -мерный объём - это просто длина (измеряемая, например, в метрах), в случае двух измерений - площадь (т.е. метры квадратные), в случае трёх измерений - привычный нам трехмерный объём (измеряемый, например, в литрах или метрах кубических), в случае четырех измерений - некий гиперобъём 4-й степени (то есть имеющий размерность  $m^4$ ) и так далее.

Теперь возьмём для наглядности, произвольную двумерную точку  $x$  с координатами  $(x_1, x_2)$ . Пусть она является центром квадрата с длиной сторон  $h$ . Тогда для любой точки  $x_i$ , принадлежащей множеству точек  $x_1, x_2, \dots, x_N$ , которые мы хотим прокластеризовать, имеющей координаты  $(x_{1i}, x_{2i})$  выполняется соотношение

$$\phi\left(\frac{x-x_i}{h}\right) = \begin{cases} 1 & |x_j - x_{ij}| \leq \frac{h}{2} \\ 0 & \text{иначе} \end{cases} \quad j=1, \dots, n$$

где  $n$  количество координат (измерений), причем это число не надо путать с  $N$  - количеством точек которые надо прокластеризовать.

Или, если стараться формулировать максимально просто:

$$\phi\left(\frac{x-x_i}{h}\right) = \begin{cases} 1 & \text{если } x_i \text{ находится в середине гиперкуба со стороной центром в } x \\ 0 & \text{ином случае} \end{cases}$$

Теперь мы можем подсчитать общее количество точек из множества точек  $x_1, x_2 \dots x_N$ , которые находятся внутри гиперкуба со стороной  $h$  и центром в точке  $x$ .

$$k = \sum_{i=1}^N \phi\left(\frac{x-x_i}{h}\right)$$

Далее мы можем разделить полученную величину на  $h^n$  - объём нашего  $n$ -мерного куба (гиперкуба), получив плотность кластеризуемых точек в гиперкубе с центром в произвольной точке  $x$ . В случае одномерного гиперкуба, то есть отрезка длины  $h$ , получим количество точек приходящихся на единицу длины.

Здесь мы остановимся и посмотрим на нашу функцию  $\phi(u)$  несколько повнимательнее. Она очень проста, но обладает несколькими не очень приятными свойствами, например, она обладает некоей симметрией, но не полностью симметрична относительно оси  $Y$  (для двумерного случая). Она так же несколько «угловата» не является везде гладкой, если говорить строго, и значит в некоторых точках у этой функции нет производной. Можно ли применить для определения плотности кластеризуемых точек в окрестности произвольной точки какие-то более подходящие классы функций? Оказывается можно. Собственно, мы подошли к концу нашей «наводящей» цепочки рассуждений. Сделаем еще несколько оговорок.

Авторы идеи, очень упрощенный вариант которой мы привели выше, Emanuel Parzen и Murray Rosenblatt, в своих рассуждениях (более математически строгих, чем наши), оперируют не с количеством точек вообще, а с количеством точек попадающих в «зону действия» нашей функции, делённой на количество всех точек  $N$ . Фактически, они оперируют с вероятностью, и деля на гиперобъём, получают плотность вероятности попадания произвольной точки в «зону действия» функции  $\phi$  (эта «зона действия» называется «окно Парзена» (Parzen window) [4] и определяется параметром  $h$ ). То есть функция, с помощью которой мы вычисляем плотность - это плотность вероятности, и она должна обладать свойствами таковой.

Формулируем более строго. Вводим непрерывную функцию плотности вероятности, тогда с этой функцией мы можем оперировать привычными нам средствами математического анализа, например найти её максимумы, которые и являются центрами уплотнений, то есть центрами кластеров. Ну а наш набор точек  $x_i, i = 1... N$  - это некое дискретное подмножество области определения функции плотности вероятности. И нам надо на основе этого набора точек построить эту функцию. Находим ее как линейную сумму функций, удовлетворяющих определённым свойствам. В англоязычной литературе по статистике/машинному обучению такие функции называются kernel - «ядро» или, если угодно, «ядерной функцией» ( $K(x)$ ). Kernel-функция обладает следующими свойствами (считаем, что точка в окрестности которой нам надо вычислить плотность вероятности, это точка начало координат): 1.  $K(x)$  «быстро», например, экспоненциально стремится к нулю по мере удаления от начала координат; 2.  $K(x)$  имеет максимум в начале координат; 3.  $K(x)$  симметрична.

Ну и для полной определённости можно сразу же и добавить, что 
$$\int_{-\infty}^{+\infty} k(x) dx = 1$$

Этим свойствам обладает функция плотности вероятности нормально распределённой случайной величины, то есть кривая Гаусса.

$$k(x) = ce^{-\frac{x^2}{2\sigma^2}}$$

Теперь мы можем для каждой точки  $x_i$  нашего набора точек определить свой kernel  $k(x - x_i)$  с центром в данной точке  $x_i$ . Выглядят точки с их kernel-ами примерно так, как показано на рисунке 1.

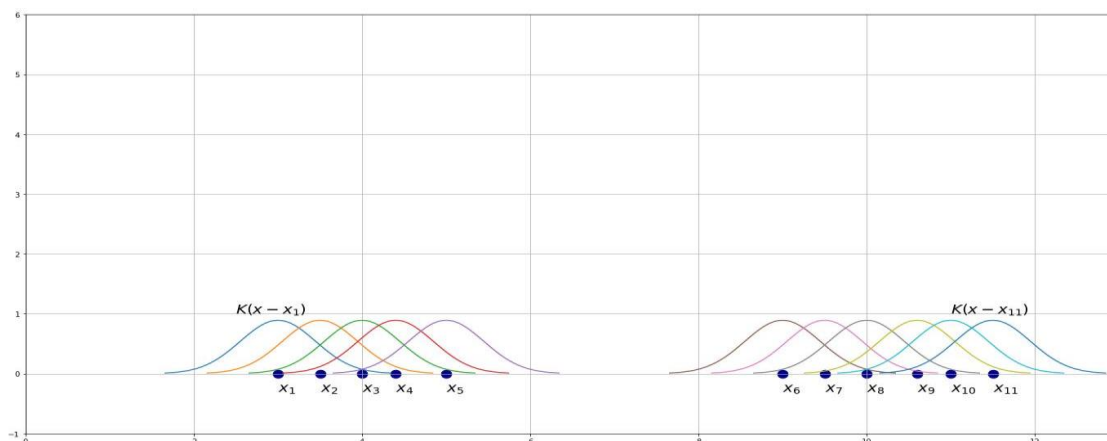


Рисунок 1. Совокупность гауссовых ядер в точках  $x_i$

Далее мы суммируем все эти kernel-ы, деля их сумму на количество  $N$  точек в нашем наборе, находя таким образом функцию плотности распределения наших точек, осуществив переход к непрерывной функции распределения плотности вероятности:

$$f(x) = \frac{1}{N} \sum_{i=1}^N k(x - x_i)$$

В принципе, мы можем думать о kernel, как о некоей «мере схожести», или «мере подобия» двух произвольных точек  $x_i$  и  $x_j$  нашего набора точек. То есть чем больше значение  $K(x_i - x_j)$ , тем больше «похожи» точки  $x_i$  и  $x_j$ .

Сделаем ещё один шаг, уточняя понятие kernel. На самом деле kernel - это, для большинства применяемых на практике kernel-ов, функция от расстояния  $\|x_i - x_j\|$  между точками:

$$k(x) = ck \left( \left\| \frac{x - x_i}{h} \right\| \right),$$

где  $c$  - нормировочная константа, вводимая для того что бы интеграл от kernel по всей области определения был равен 1. В англоязычной литературе по статистике/машинному обучению  $h$  это bandwidth или window bandwidth - "ширина окна". Имеется ввиду ширина Парзеновского окна.

Для нас важно, что величина  $h$  характеризует «радиус влияния» точки  $x_i$ . От выбора величины этого параметра зависит ключевой момент применения MeanShift алгоритма, о котором будет сказано дальше.

Итак, введя расстояние между точками нашего набора точек мы фактически дали описание не только одномерного случая, но и многомерного. Хотя, если под расстоянием подразумевается обычное евклидово расстояние, то мы тогда ограничиваем себя так называемыми radially symmetric kernels, то есть «кругосимметричными ядрами» (в начале наших рассуждений уже была формулировка «вокруг начала координат»). Впрочем, для обобщения на многомерный случай радиальность kernel-а значения не имеет. Но для упрощения рассуждений и вычислений будем оперировать именно с radially symmetric kernel.

Итак, в общем случае, в пространстве с евклидовой метрикой, такой kernel имеет вид:

$$k(x) = ck \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)$$

Соответственно, функция плотности вероятности имеет вид:

$$k(x) = \frac{1}{N} \sum_{i=1}^N ck \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)$$

В вышеуказанной англоязычной литературе данный метод аппроксимации плотности вероятности на конечном множестве точек называется Kernel Smoothing (ядерное сглаживание).

Теперь мы можем найти локальные максимумы функции плотности вероятностей воспользовавшись gradient ascent методом, то есть методом градиентного восхождения, который на самом деле ничем не отличается от метода градиентного спуска за исключением знака с которым берётся градиент. Итак, в общем виде, алгоритм нахождения всех максимумов функции плотности выглядит так:

1. Для всех  $i, \dots, N$  повторять:

$$x \leftarrow x_i + \nabla f(x) = x_i + \frac{1}{N} \nabla \sum_{i=1}^N ck \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)$$

2. До тех пор, пока  $x$  перестанет изменяться.

$\nabla$  - набла-оператор - это векторная сумма первых частных производных, а в случае одного измерения, просто первая производная.

Далее, специальным образом находится градиент  $\nabla f(x)$ , с помощью введения обозначения  $g(t) = k'(t)$ .

При этом разность между  $x$  и средневзвешенным по  $g(\bullet)$  от точек, которые лежат «близко» от  $x$  (то есть тех, которые попадают в window точки  $x$ ), и называется mean shift вектором. Обозначим его через  $M(x)$ . Очевидно, что он имеет то же направление, что и градиент. Воспользовавшись  $M(x)$ , окончательно запишем выражение для каждого последующего  $x$  в нашем алгоритме как:

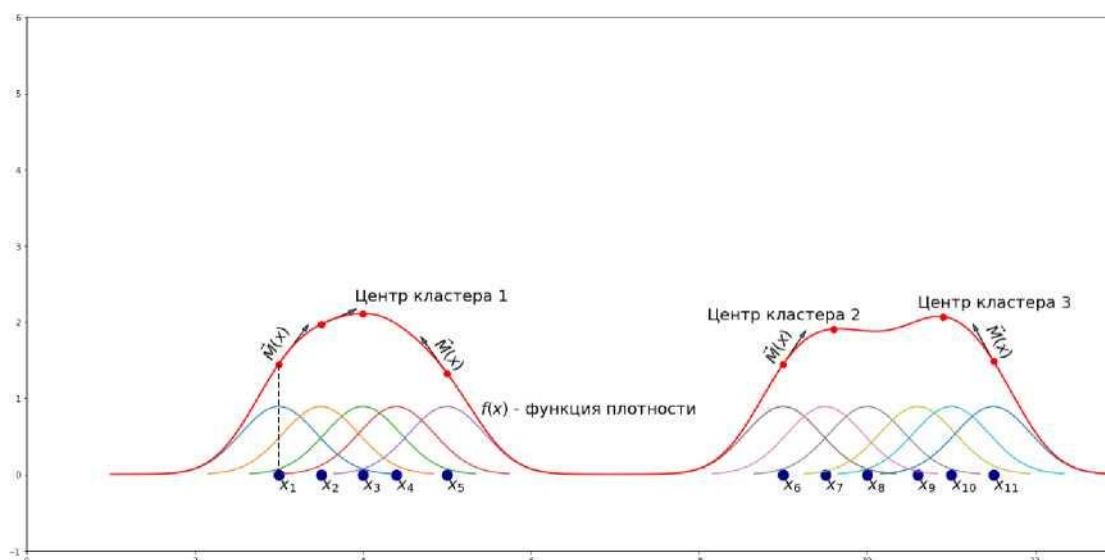
$$x \leftarrow x + M(x) = x + \left( \frac{\sum_{i=1}^N x_i g \left( \frac{\|x - x_i\|^2}{h^2} \right)}{\sum_{i=1}^N g \left( \frac{\|x - x_i\|^2}{h^2} \right)} - x \right) = \frac{\sum_{i=1}^N x_i g \left( \frac{\|x - x_i\|^2}{h^2} \right)}{\sum_{i=1}^N g \left( \frac{\|x - x_i\|^2}{h^2} \right)}$$

Чаще всего в качестве  $k(x)$  используется уже упоминавшаяся функция Гаусса. В этом случае  $h$  это среднеквадратическое отклонение (обычно обозначается как  $\sigma$ ). О подборе величины параметра  $h$  (bandwidth) следует сказать следующее. Если мы его выберем слишком малым, то у нас получится множество кластеров, и общая картина получится слишком «зашумленной». Если же задать его слишком большим, то картина получится слишком сглаженной, с малым количеством кластеров, из-за чего мы рискуем не увидеть важные закономерности. Сделаем уточнение по поводу «близко-далеко» и «зоны влияния». Конечно, для Гауссова ядра зона влияния - это вся числовая ось. Но на практике точки лежащие например, за пределами широкоизвестных «трёх сигма» можно не учитывать. Для некоторых же типов kernel-ов «зона действия» имеет чёткую границу (например, для функции  $\phi$ , которую мы использовали выше).

Важнейшее преимущество Mean Shift алгоритма в том, что он не требует заранее задавать количество кластеров (в отличие например от наиболее популярного алгоритма «обучения без учителя» K-Means), то есть не заставляет заранее нас делать какие либо предположения (а скорее, догадки) о количестве или расположении кластеров.

В простейшем, одномерном случае мы можем проиллюстрировать работу алгоритма, дополняя предыдущий рисунок 1 с набором точек и их kernel-ами, следующим образом (рис.2).





*Рисунок 2. Суперпозиция гауссовых ядер в точках  $x_i$ , позволяющая отыскивать локальные максимумы суперпозированной функции, то есть центры кластеров*

В кратце Mean Shift алгоритм можно описать следующим образом:

1. Строим на множестве точек, которые надо прокластеризовать, непрерывную функцию плотности вероятности, воспользовавшись ядерным сглаживанием (kernel smoothing).

2. Используя модифицированный метод градиентного восхождения, в котором градиент от функции плотности вероятности, полученной на предыдущем шаге, заменен на вектор сдвига от данной точки к средневзвешенному значению координат точек попадающих в Парзеновское окно данной точки (Mean Shift vector), приходим из каждой точки набора точек, которые нам необходимо прокластеризовать, в ближайший к точке локальный максимум.

3. Точки для которых локальные максимумы совпадают, объявляем принадлежащими к одному кластеру.

Создаем экземпляр класса MeanShift, задавая bandwidth. ScikitLearn предоставляет возможность во многих случаях вычислить его при помощи функции estimate\_bandwidth, избавив нас от необходимости его угадывать. Далее вызываем стандартный для всех классов ScikitLearn, реализующих методы машинного обучения, метод fit, передавая ему массив обучающих значений. После обращаясь уже к «обученному» (произведшему кластеризацию) экземпляру класса выводим количество получившихся кластеров, их получилось 4.

Выводим (используя библиотеку Seaborn) кросскорреляционную матрицу. Кроме значения коэффициента корреляции степень корреляции выделяется цветом (рис. 3).

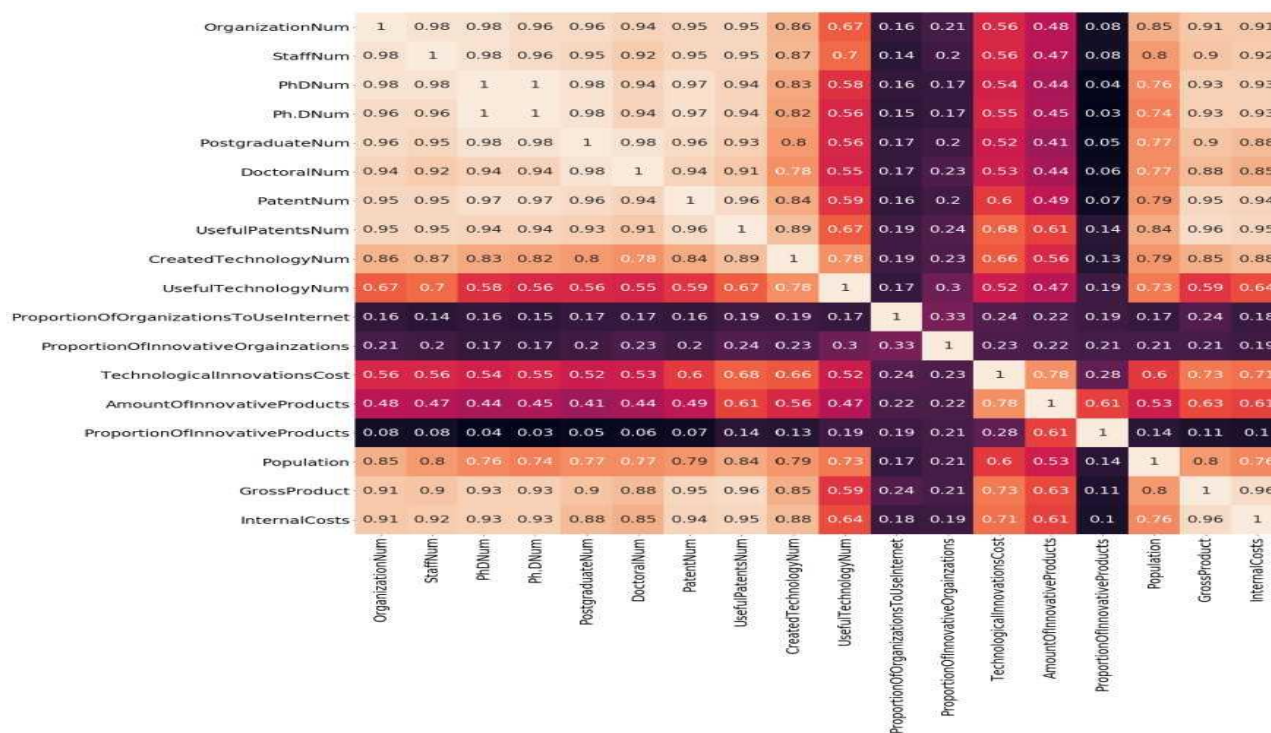


Рисунок 3. Кросскорреляционная матрица для 18 индикаторов задачи кластеризации

Для упрощения восприятия, выводим соответствия между названиями (англоязычными) столбцов в датафрейме и соответствующими русскоязычными названиями показателей. Так же выводим и соответствующие номера столбцов. Нумерация начинается с нуля, как это принято в большинстве языков программирования до 17 и соответствует расположению всех 18 индикаторов на кросскорреляционной матрице.

Для понимания полученной в результате кластеризации информации нам необходимо эту информацию отобразить графически, а для этого необходимо выделить из набора всех показателей наиболее «важные» или значимые. Мы можем для этого использовать полученную шагом ранее кросскорреляционную матрицу выбрав наименее коррелирующие пары индикаторов. Но этот подход выявляет только линейные зависимости, и к тому же нам придется самим как-то писать дополнительный программный код для автоматизации этого выбора (что довольно громоздко), либо выбирать пары вручную, что трудно и затратно по времени. Попробуем применить другой подход. Воспользуемся тем, что некоторые методы «обучения с учителем» (supervised learning) могут ранжировать признаки (features, столбцы) по степени важности. В частности таким свойством обладают методы основанные на объединении (ensemble) деревьев решений, например Random Forest («Лес случайных деревьев» Алгоритм предложен Лео Брейманом и Адель Катлером: [3], [5], [6], [7]). Имеются ввиду decision trees «решающие деревья» или «деревья решений» [3] [6], [7] [8]. В каждом узле такого дерева происходит вычисление энтропии по Шеннону, либо значения «критерия загрязнённости» Джинни (Gini impurity) [3] [7], [9], что автоматически даст нам так же выявление нелинейных зависимостей. А разметка по классам, необходимая для применения методов supervised learning, у нас уже есть - принадлежность к тому или иному кластеру мы можем трактовать, как принадлежность к соответствующему классу. Создаем копию датафрейма и отдельное поле (столбец), которое и будет содержать метку класса (номер кластера).

Определяем вспомогательную функцию, которая понадобится для определения лучшего сочетания параметров выбранного метода supervised learning (Random Forest-a)

Создаём экземпляр стандартного ScikitLearn класса RandomForestClassifier и производим определение наилучшего сочетания значений параметров с помощью

RandomForestClassifier-а посредством простого поиска по сетке значений параметров. Определяем допустимые диапазоны значений различных параметров Random Forest-а, для дальнейшего применения поиска наилучшего сочетания значений параметров по сетке значений. Создаём экземпляр класса GridSearchCV, реализующего алгоритм поиска по сетке.

Производим поиск по сетке значений параметров и создаём экземпляр randomforestclassifier-а, подставляя найденный наилучший набор значений параметров.

Запускаем процесс обучения классификатора и определяем степени важности индикаторов (features) инновационности (номера соответствующих столбцов) и выводим диаграмму в графическом виде представляющую степени важности индикаторов (в виде высоты столбцов) (рис.4). Эти ранжированные индикаторы с коэффициентами весомости приведены ниже:

1. Индикатор 10 (0.188041)

Удельный вес организаций, использовавших интернет по субъектам российской федерации (в процентах от общего числа обследованных организаций соответствующего субъекта российской федерации);

2. Индикатор 11 (0.167093)

Инновационная активность организаций (удельный вес организаций, осуществлявших технологические, организационные, маркетинговые инновации в отчетном году, в общем числе обследованных организаций ( %);

3. Индикатор 9 (0.082238)

Используемые передовые производственные технологии по субъектам российской федерации;

4. Индикатор 1 (0.067305)

Численность персонала без ученых степеней, занятых НИОКР;

5. Индикатор 14 (0.056109)

Удельный вес инновационных товаров, работ, услуг в общем объеме отгруженных товаров, выполненных работ, услуг, по субъектам российской федерации (%);

6. Индикатор 13 (0.055939)

Объем инновационных товаров, работ, услуг (миллионов рублей);

7. Индикатор 8 (0.052259)

Разработанные передовые производственные технологии по субъектам Российской федерации;

8. Индикатор 7 (0.051805)

Количество патентов, выданных на полезные модели;

9. Индикатор 0 (0.043066)

Организации, выполняющие научные исследования и разработки;

10. Индикатор 16 (0.042287)

Валовой региональный продукт по субъектам Российской Федерации (в текущих ценах; миллионов рублей);

11. Индикатор 4 (0.032444)

Численность аспирантов по субъектам Российской Федерации – человек;

12. Индикатор 15 (0.031357)

Численность населения по субъектам Российской Федерации (оценка на конец года; тысяч человек);

13. Индикатор 6 (0.028565)

Количество патентов, выданных на изобретения;

14. Индикатор 12 (0.028055)

Затраты на технологические инновации организаций (руб);

15. Индикатор 17 (0.024900)

Внутренние затраты на научные исследования и разработки миллионов рублей;

16. Индикатор 3 (0.020874)

Численность исследователей, имеющих ученую степень доктора наук, по субъектам Российской Федерации показатель – человек;

17. Индикатор 2 (0.020040)

Численность исследователей, имеющих ученую степень кандидата, по субъектам Российской Федерации - человек;

18. Индикатор 5 (0.007623)

Численность докторантов по субъектам Российской Федерации –человек.



*Рисунок 4. Ранжирование индикаторов по степени их важности*

Из рисунка видим, что самым важным (весомым) индикатором оказался индикатор с номером десять (считая от нуля) - «удельный вес организаций использовавших интернет по субъектам Российской Федерации». В целом, было проведено множество запусков RandomForest для выявления закономерностей в упорядочивании показателей по значимости. В каждом узле дерева решений (каждого из деревьев решений, которые строит Random Forest) сплит (англ. split), то есть разбиение по значениям очередного индикатора, происходит среди показателей случайно отобранных из всего множества индикаторов, почему алгоритм и называется Random Forest. По этой причине от запуска к запуску, в общем случае, упорядочение меняется. Мы можем, конечно, изменить это поведение алгоритма посредством задания параметра `random_state`, но нас интересуют именно статистические закономерности. Индикаторы 10 и 11 входят в пятёрку самых значимых почти при каждом перезапуске (по крайней мере один из них, чаще - оба), причём обычно 10-й индикатор оказывается на 1-ом месте. Очень часто входят в пятёрку индикаторы 16, 13, 1, 9, 14. Мы остановились на варианте, когда в пятёрке самых значимых присутствуют 10-й и 14-й показатели, так как 14-й к тому же весьма слабо коррелирует (линейно) со всеми остальными индикаторами, как видно из кросскорреляционной матрицы (рис.3). Так же наличие 14-го индикатора на scatter plot диаграммах весьма чётко иллюстрирует один из выводов сделанных далее.

Далее нам надо отобразить кластеризованные регионы на двумерных проекциях парных сочетаний соответствующих индикаторов. Для этого мы и выбирали наиболее зна-

чимые индикаторы, чтобы сократить число выводимых двумерных проекций. Задача графического отображения кластеризованных регионов на проекциях требует указания множества различных параметров отображения, и является довольно громоздкой. Для её решения был создан класс ClustersDrawer, исходный код (текст) которого вынесен в отдельный файл clusters\_drawer.py

Среди задаваемых параметров следует выделить ms - обученный MeanShift кластеризатор, содержащий информацию о кластерах (принадлежность того или иного региона к тому или иному кластеру), и список significant\_features - номера наиболее важных индикаторов.

Для облегчения восприятия, определены попарные сочетания названий пяти наиболее значимых индикаторов, ниже показаны первая и последняя пары:

0. удельный вес организаций, использовавших Интернет по субъектам Российской Федерации (в процентах от общего числа обследованных организаций соответствующего субъекта Российской Федерации), %;

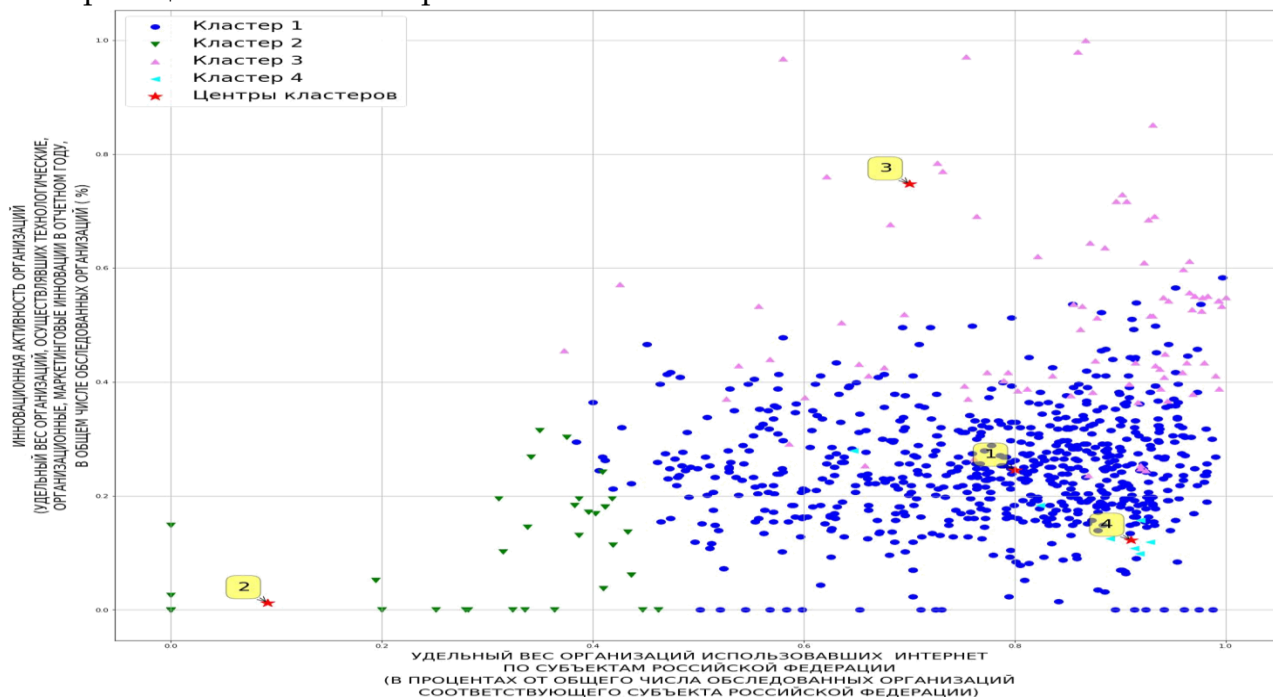
0. количество патентов, выданных на полезные модели, n;

.....

9. удельный вес инновационных товаров, работ, услуг в общем объеме товаров, работ, услуг по субъектам Российской Федерации, %;

9. численность персонала без ученых степеней, занятых НИОКР, человек.

На рисунке 5 показаны проекции многомерных точек на плоскость “удельный вес организаций, использовавших интернет” – “инновационная активность организаций”. Остальные проекции показаны в Приложении 1.



**Рисунок 5.** Кластеризованные проекции многомерных объектов на плоскость “удельный вес организаций использовавших интернет” – “инновационная активность организаций”

Сразу стоит указать, что на каждом изображении (двумерной проекции) содержится полный временной (ударение на «о») слепок каждого кластера. То есть каждый регион на изображении встречается несколько раз, (столько раз сколько лет во временном диапазоне на протяжении которого мы кластеризуем), причём он может в общем случае быть обозначен разными значками с разным цветом, обозначающими принадлежность к разным кластерам (может с течением времени «путешествовать» по кластерам). Далее можно попробовать сделать некие общие выводы глядя на построенные нами scatter plot

диаграммы, и попробовать определить критерии принадлежности к каждому кластеру (или иначе - попробовать на понятном языке как-то назвать полученные кластеры). В качестве примера, подробнее разберём рисунок, на котором изображена зависимость «степени инновационной активности организаций» в зависимости от «степени интернетизации».

Глядя на изображение можно сделать вывод, что большая часть регионов на протяжении всего срока наблюдения попадает в кластер 1, для которого довольно высокие значения независимой переменной (степени интернетизации) не приводят тем не менее к высоким значениям зависимой (ну или лучше сказать «коррелирующей, причём возможно нелинейно») переменной (степени инновационной активности организаций).

Кластер 2 можно охарактеризовать тем, что низкие значения независимой переменной соответствуют низким значениям зависимой.

Кластер 3 можно охарактеризовать тем, что для него высокие значения независимой переменной соответствует высоким значениям зависимой.

Про кластер 4 из анализа первого изображения пока ничего определённого сказать нельзя. На первый взгляд, он не отличается от кластера 1.

Анализ последующих изображений подтверждает выводы о кластерах 1, 2, 3, и проясняет характеристики кластера 4. Итак:

1. *Кластер 1.* Можно охарактеризовать понятием «среднячки», или «эффективность от невысокой до средней», или «не очень высокая инновационная отдача на относительно (или даже абсолютно) высокие вложения».

2. *Кластер 2.* Можно охарактеризовать понятием «отстающие», или «малая инновационная отдача в ответ на малые вложения».

3. *Кластер 3.* «Лидеры». Характеризуются высокой инновационной отдачей в ответ на высокие вложения.

4. *Кластер 4.* Можно охарактеризовать понятием «аномальная эффективность». По некоторым параметрам регионы ведут себя как представители класса 1, но по некоторым показывают высокую инновационную отдачу в ответ на низкие вложения (что отчётливо видно на диаграммах с 14-м индикатором: «инновационная активность организаций»). Это не исключает того, что по некоторым индикаторам в 4-ом кластере, возможно, как раз аномально низкое значение одного из индикаторов в сочетании с высоким значением другого. Впрочем, такая ситуация не исключена и в других кластерах.

Здесь и далее под инновационной отдачей понимается влияние входным переменных, характеризующих инновационный потенциал региона (охват населения Интернетом, кадровый научный потенциал, расходы на НИОКР и др.), на результирующие переменные (количество созданных инновационных предприятий, инновационных технологий, количество выданных патентов на изобретения и др.).

Выводим списки регионов сгруппированные в кластеры по годам. Для этого добавляем в главный датафрейм, с которым мы оперируем, информацию о принадлежности региона к кластеру в виде столбца "cluster\_label". Разделяем главный датафрейм на список датафреймов, в котором каждый элемент списка - отдельный кластер.

Определяем функцию для разбиения кластера (датафрейма, в котором регионы сгруппированы по кластерам) по годам. Определяем временной диапазон. Разбиваем кластеры по годам. Из полученного разбиения кластеров по годам можно сделать несколько выводов:

1. Не во все годы в кластер 4 («Аномальная эффективность») попадает хотя бы один регион. Более того, регионы начинают попадать в кластер 4, начиная с середины срока наблюдения, и в особенности ближе к концу срока.

2. Кластер 2 («Отстающие») со временем исчезает («рассасывается»), регионы перемещаются из него, в основном, в кластер 1 («Среднячки»).

3. Большая часть регионов надежно остаются в своих кластерах, но есть и регионы перемещающиеся в другие кластера (например, регионы из разряда «Отстающие» перемещающиеся в «Середнячки») и регионы «бродящие» по различным кластерам, например, Самарская область, или Архангельская область, которые из «Отстающих» переместилась в «Аномальные».

4. Большая часть регионов со временем сосредоточилась в кластере 1.

Далее необходимо решить задачу отображения информации о кластеризованных регионах на карте административных единиц Российской Федерации. Для этого загружаем значения кодов регионов. Загружаем карту административных единиц РФ с сопутствующей информацией. Карты в формате Shapefile находятся в свободном доступе: <https://gadm.org/data.html>. Для работы с картами в данном формате используется Python библиотека с таким же названием.

Из загруженных данных читаем информацию о регионах (содержащую, в частности, названия регионов), и собственно, само геометрическое описание регионов (в виде неких многоугольников).

Дальше определено несколько вспомогательных функций/участков кода, в основном, предназначенных для сопоставления названий регионов (административных единиц) взятых из различных источников: данных о инновационной активности, кодов регионов, географической информации. Из каждого названия выделяется основная (значимая) часть. Многие регионы в различных источниках пишутся по-разному, применяются различные аббревиатуры, и т.д., поэтому некоторые названия пришлось сопоставлять вручную. Имена (названия) определяемых функций и переменных даны с учетом того, что имя должно сообщать максимальную информацию о смысле данной функции (переменной). Так же, с помощью определённых далее вспомогательных функций, производилось сопоставление регионов (их индексов в соответствующих списках) из различных источников, на основе ранее выделенных главных частей названий, что нужно для отрисовки регионов с их кодами на карте.

Отдельная задача состояла в отрисовке кода региона на карте, так чтобы код попадал «внутри» границ региона. Стандартных средств для этого в библиотеках Python нет. Задача решена в некоторых проприетарных пакетах, но налаживание взаимодействия с ними (хотя бы в виде уже готовых карт) может занять весьма неопределенное время. Поэтому была сделана попытка решить некий простейший вариант данной задачи, с возможностью «ручной подстройки» координат и размера надписи. В общем виде решение данной задачи - разместить объект, ограниченный неким прямоугольником «где-то внутри» произвольного, и в общем случае, невыпуклого многоугольника «ближе всего» к некоторому наиболее удобному для человеческого глаза «центру восприятия» многоугольника (смасштабировав к тому же размещаемый объект) превосходит как минимум в несколько раз по трудо- и время затратам данную работу. Как показали проведенные эксперименты, достаточно неплохие результаты во многих случаях дает просто размещение надписи в центре ограничивающего регион прямоугольника (bbox в терминологии shapefiles). Размещение в центре масс многоугольника дает худший результат, в особенности, если регион имеет извилистую береговую линию (тогда центр масс сильно к ней смещён). О регионах, состоящих из нескольких островов и говорить не приходится, для них поможет только установка надписи «вручную», как, скорее всего, и сделано в проприетарном ПО. Следует оговорить, что далее нами нарушено одно из основных правил написания хорошего программного кода: введено некоторое количество не именованных «магических» чисел (размер шрифта), но код у нас одноразовый, и дальнейшей поддержки и развития не предполагает, поэтому вряд ли нарушение этого правила принесёт нам большой вред.

Определяем таблицу кодов регионов, для которых координаты (и возможно, размер) надписи будем устанавливать вручную. Определяем функцию (довольно громоздкую и сложную) отрисовывающую карту регионов. Выводим названия административных единиц вместе с их кодами.

Вызываем функцию `draw_clustred_rregions` для каждого года, отрисовывая регионы принадлежащие к данному кластеру в данном году. В Приложении 2 приведены все 10 картосхем кластеризованных регионов РФ по их инновационной активности вместе с разбиением их на кластеры с кодировкой регионов. Ниже в качестве примера приведем картосхему за 2014 год с перечнем кластеризованных регионов (рис. 6, табл. 2).



Рисунок 6. Картосхема кластеризации регионов РФ, 2014 год

Таблица 2. Группировка регионов по кластерам, 2014 год

**Кластер: 1** Белгородская область, Брянская область, Владимирская область, Воронежская область, Ивановская область, Калужская область, Костромская область, Курская область, Орловская область, Рязанская область, Смоленская область, Тамбовская область, Тверская область, Тульская область, Ярославская область, Республика Карелия, Республика Коми, Архангельская область, Ненецкий автономный округ, Вологодская область, Калининградская область, Ленинградская область, Мурманская область, Новгородская область, Псковская область, Республика Адыгея, Республика Калмыкия, Краснодарский край, Астраханская область, Волгоградская область, Ростовская область, Республика Дагестан, Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия, Чеченская Республика, Ставропольский край, Республика Башкортостан, Республика Марий Эл, Удмуртская Республика, Пермский край, Кировская область, Оренбургская область, Пензенская область, Самарская область, Саратовская область, Ульяновская область, Курганская область, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ, Ямало-Ненецкий автономный округ, Челябинская область, Республика Алтай, Республика Бурятия, Республика Тыва, Республика Хакасия, Алтайский край, Забайкальский край, Красноярский край, Иркутская область, Кемеровская область, Новосибирская область, Омская область, Томская область, Республика Саха (Якутия), Камчатский край, Приморский край, Хабаровский край, Амурская область, Магаданская область, Еврейская автономная область

**Кластер: 2**



**Кластер: 3** Липецкая область, Московская область, Москва, Санкт-Петербург, Республика Мордовия, Республика Татарстан, Чувашская Республика, Нижегородская область, Чукотский автономный округ

**Кластер: 4** Сахалинская область

### Заключение

Рассматривая все 10 картосхем (Прил.2), мы можем сделать ещё несколько выводов:

1. Большая часть «Отстающих» или «Аномальных» регионов отстоят далеко от административного центра - Московского региона. Отстающие - в основном некоторые регионы Крайнего Севера, либо же - республики Кавказа. Аномальные - Сибирь, Дальний восток.

2. Большая часть «Лидеров» - Центральная Россия, Урал, Юг Поволжья: Москва, Свердловская область, Татарстан, то есть наиболее богатые/населённые административные единицы Российской Федерации.

Так же стоит заметить, что в ходе работы выявлено несколько требующих дальнейшего изучения моментов. Прояснение причин перехода регионов из кластера в кластер, дополнительное исследование регионов, относящихся к кластеру 4 («аномальной» инновационной активностью), имеет и теоретический и, возможно, практический смысл, но явно выходит за рамки данного исследования.

### Список литературы

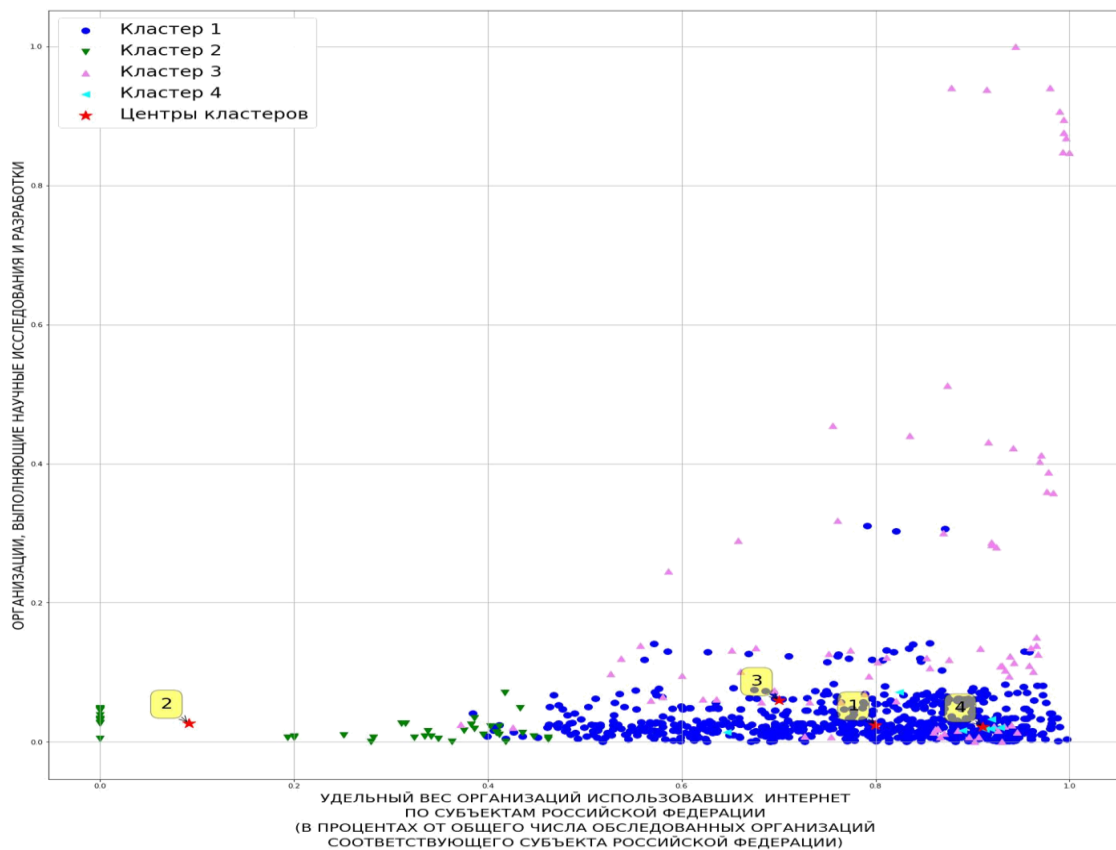
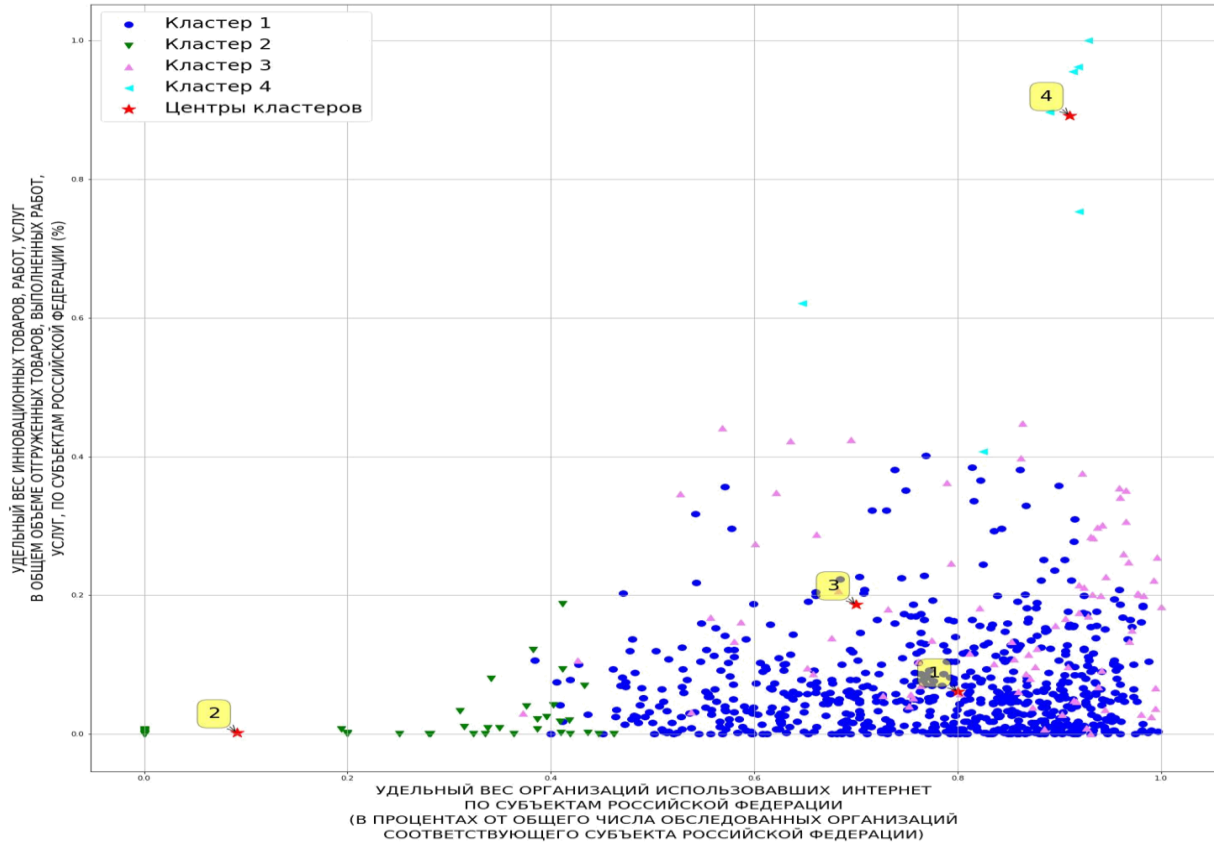
1. Московкин В.М., Сизъунго Муненге, Украинский П.А. Картографическая кластеризация инновационной инфраструктуры и инновационной активности регионов России // Оригинальные исследования: науч.-практ. электрон. журн. - 2019. - №1. - С. 57-94.
2. Saravanan Thirumuruganathan. Introduction To Mean Shift Algorithm. Неформальное введение в McanShift алгоритм с ссылками для дальнейшего изучения: режим доступа <https://saravananthirumuruganathan.wordpress.com/about/>
3. Документация библиотеки SeikitLcarn. Содержит так же огромное количество теоретических сведений и ссылок для дальнейших) изучения: режим доступа <https://scikit-learn.org/stable/documentation.html>
4. Sebastian Raschka. Kernel density estimation via the Parzen-Rosenblatt window method. Очень понятное объяснение того что такое Парзеновское окно и что такое kernel-функция: режим доступа [http://sebastianraschka.com/Articles/2014\\_kernel\\_density\\_est.html](http://sebastianraschka.com/Articles/2014_kernel_density_est.html)
5. Leo Breiman and Adele Cutler. Random Forests. Описание Random Forest алгоритма от его авторов: режим доступа [https://www.stat.berkeley.edu/~breiman/RandomForests/cc\\_home.htm](https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm)
6. Trevor Hastie Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning. Фундаментальное введение в методы машинного обучения. Наиболее известная книга в этой области: режим доступа <https://web.stanford.edu/~hastie/ElemStatLearn/>
7. Sebastian Raschka. Python Machine Learning. Практическое введение в машинное обучение с изложением: режим доступа <https://www.amazon.com/Python-Machine-Learning-Sebastian-Raschka/dp/1783555130>
8. Decision tree learning. Статья о деревьях решений в машинном обучении: режим доступа [https://en.wikipedia.org/wiki/Decision\\_tree\\_learning](https://en.wikipedia.org/wiki/Decision_tree_learning)
9. Gini Impurity vs Entropy. Вопрос на Stackсhange с краткими и ясными ответами и всеми необходимыми ссылками: режим доступа

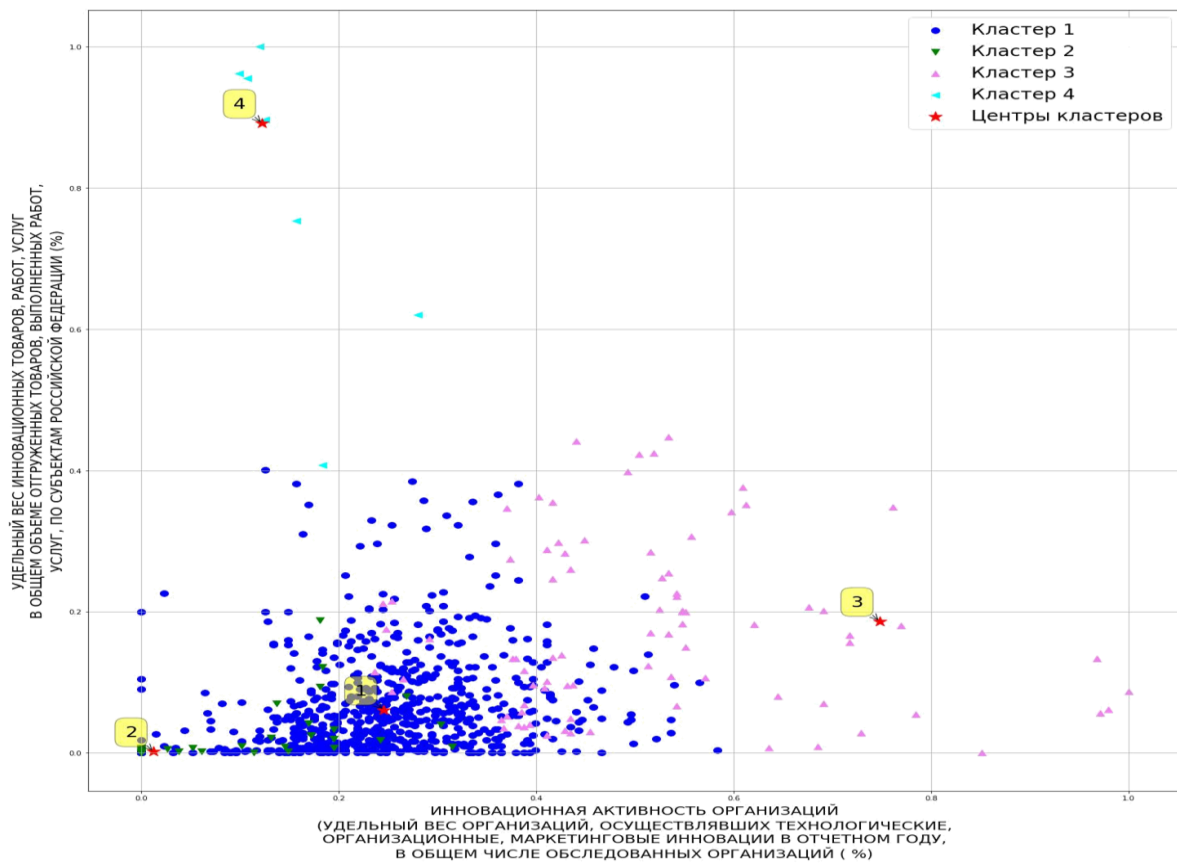
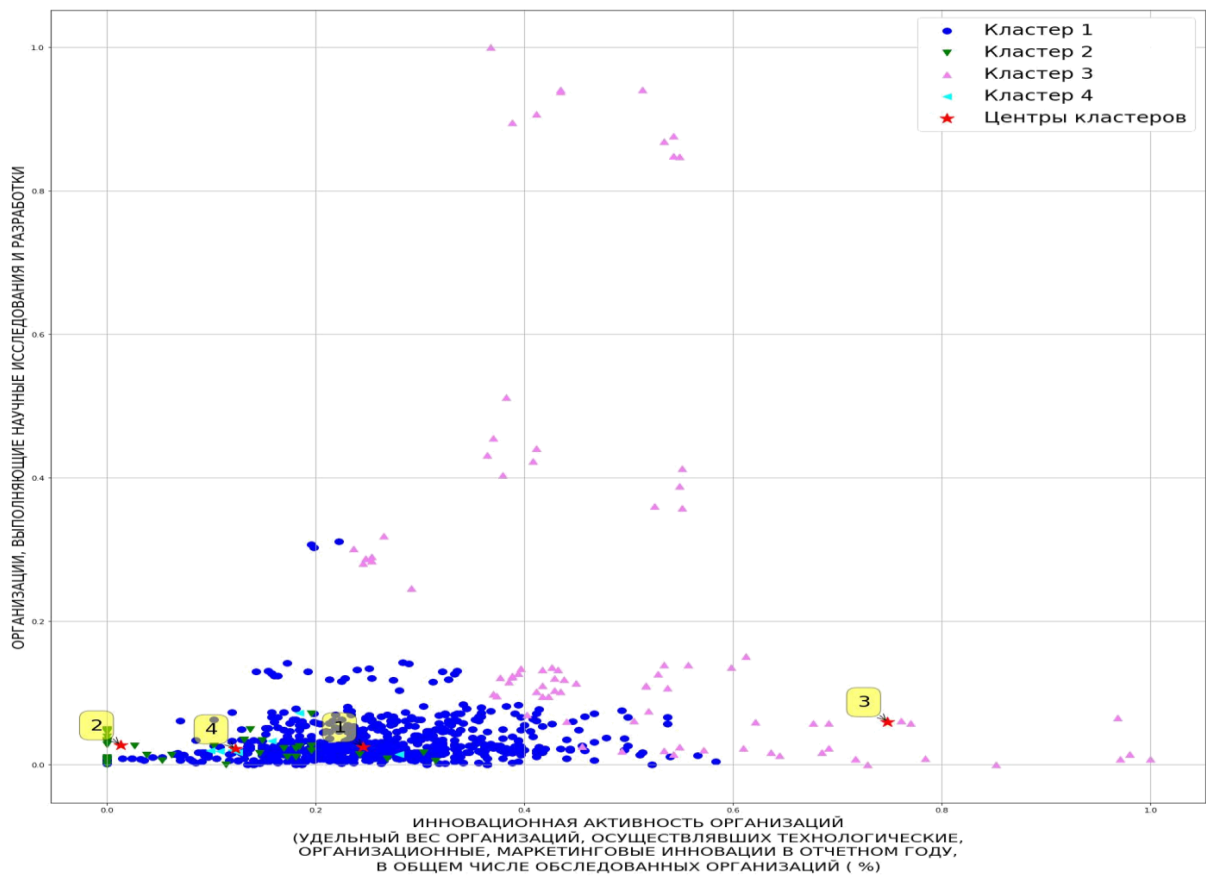
<https://datascience.stackexchange.com/questions/10228/when-should-i-use-gini-impurity-as-opposed-to-information-gain>

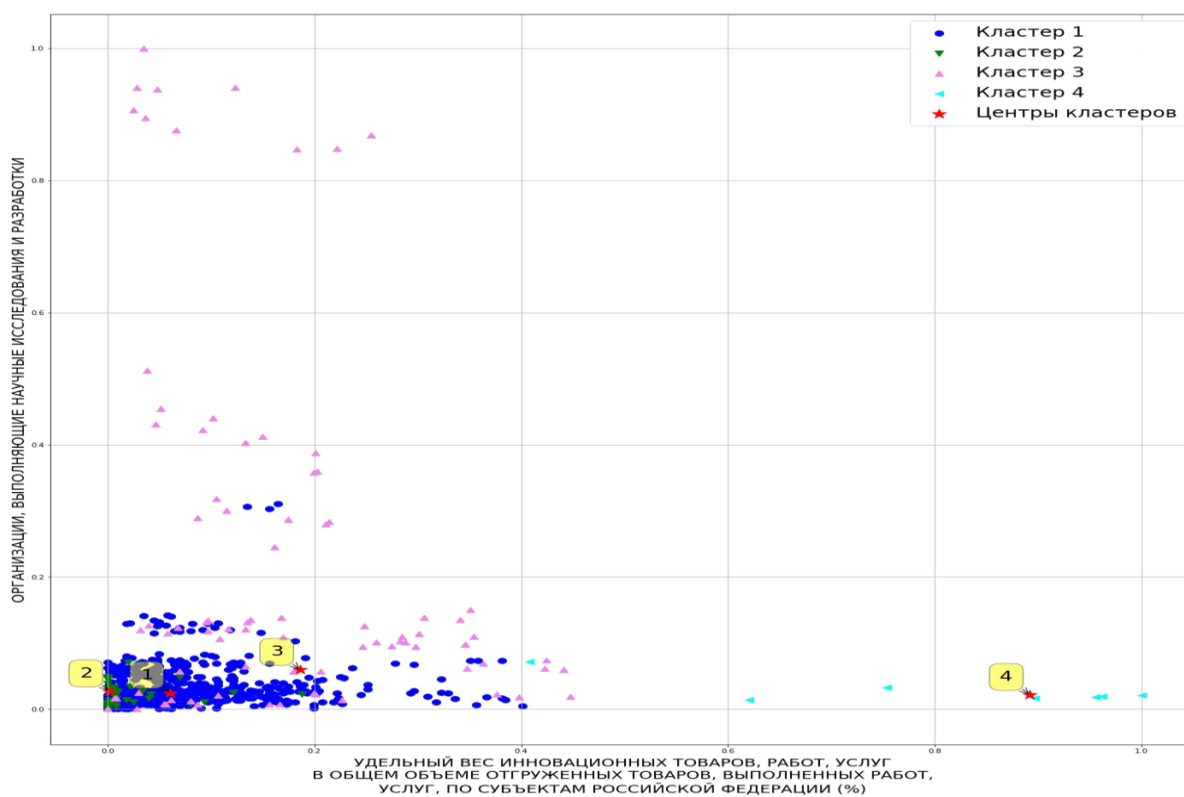
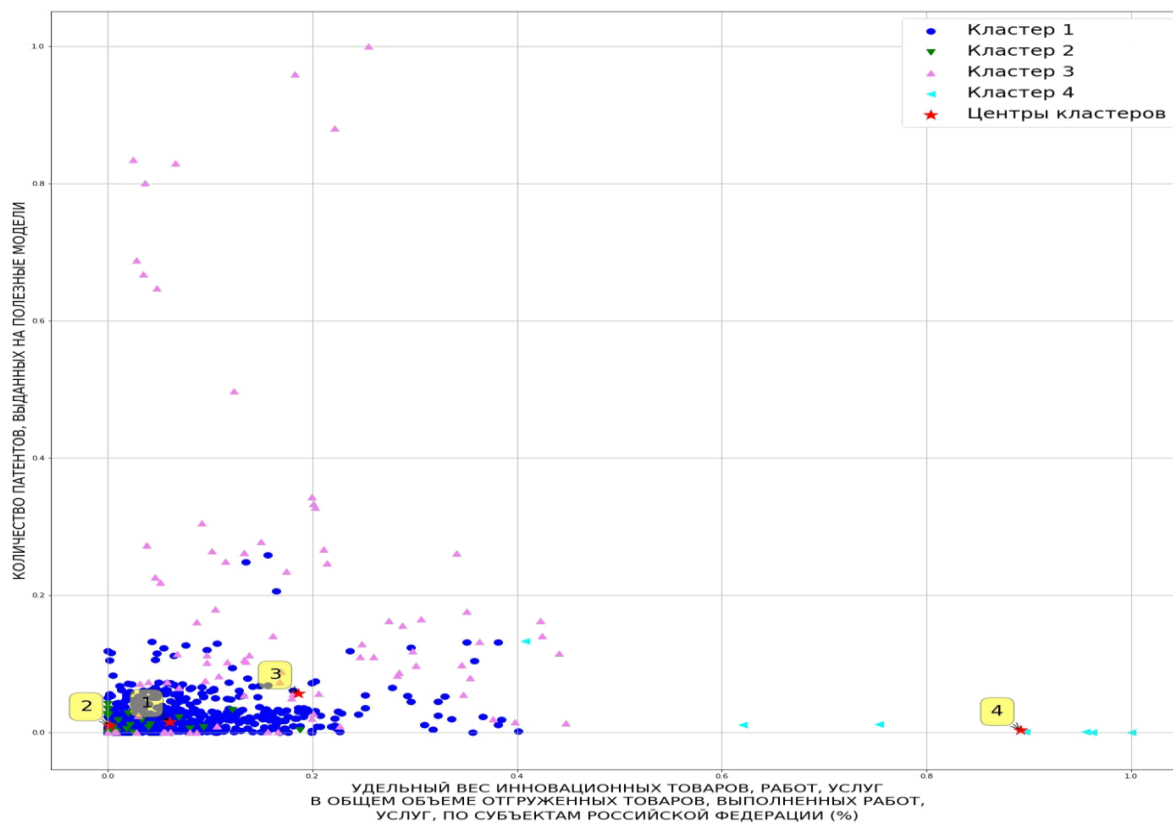
### References

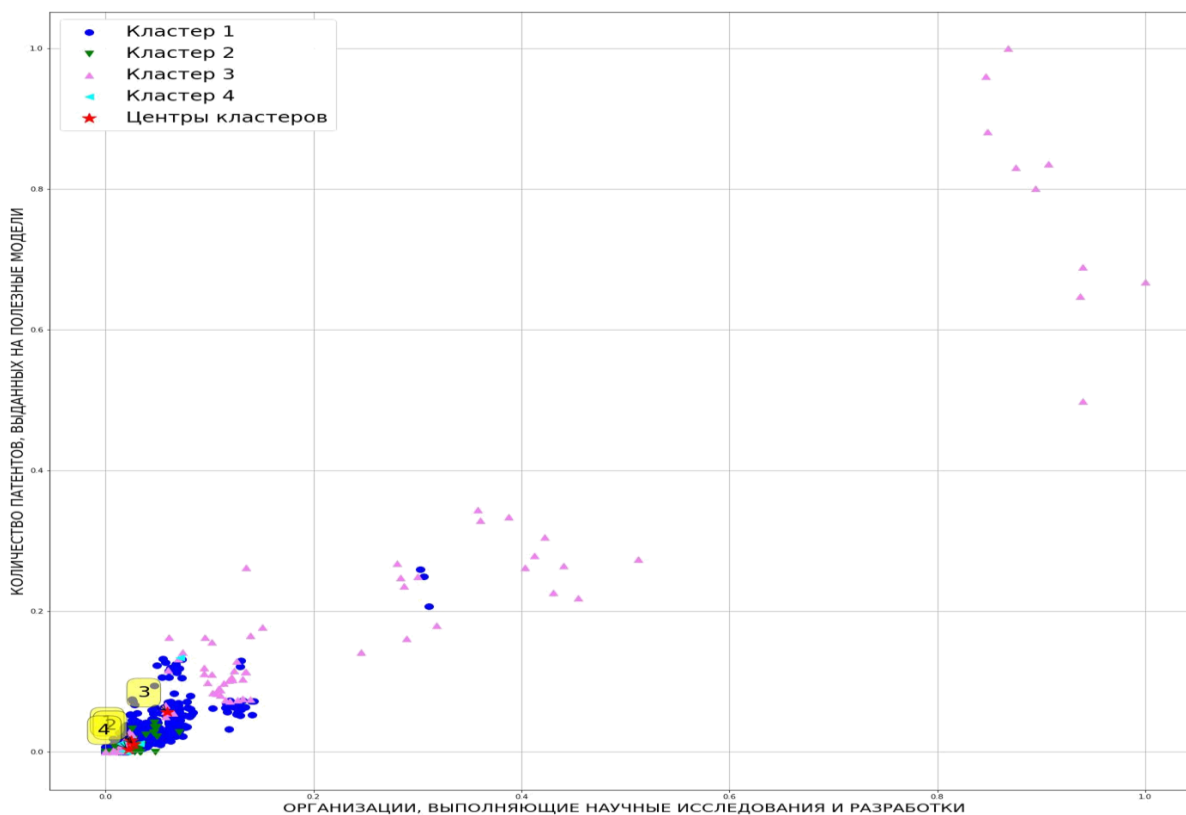
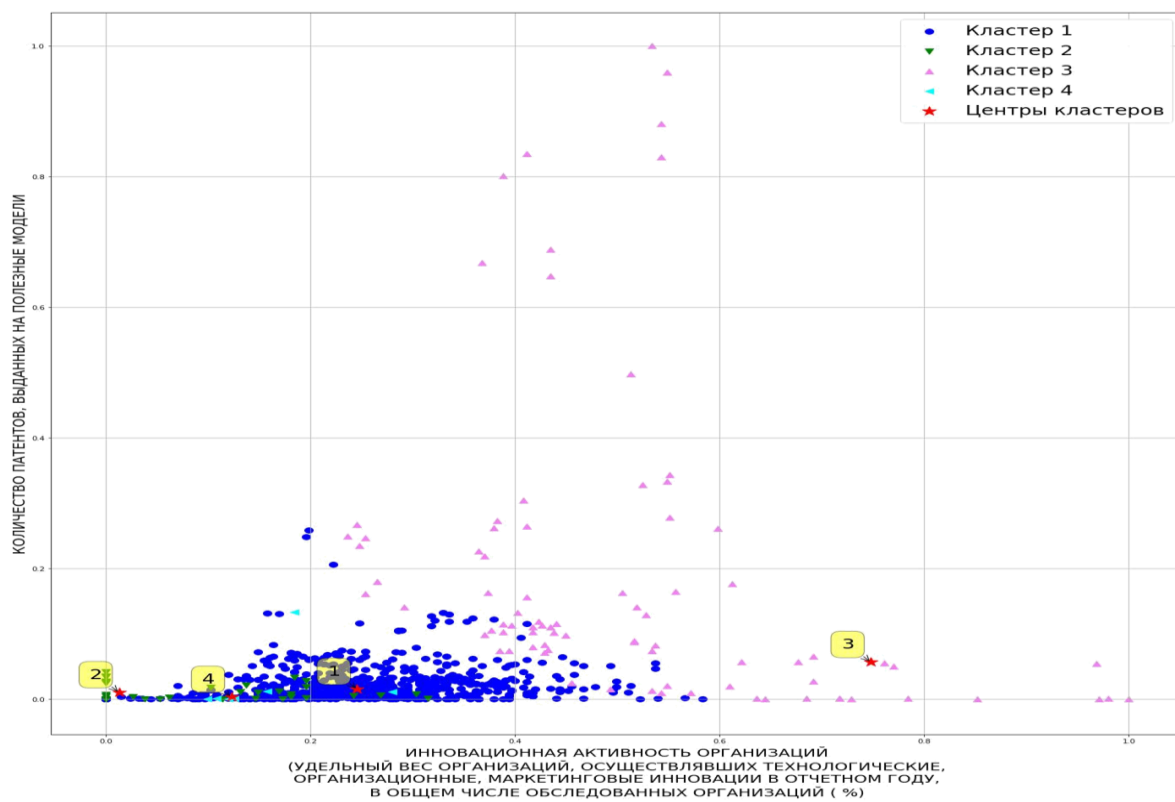
1. Moskovkin V.M., Sisungo Munenge, Ukrainian P.A. Cartographic clustering of the innovation infrastructure and innovation activity of the regions of Russia // Original research: scientific and practical works. electron. journals - 2019. - №1. - С. 57-94.
2. Saravanan Thirumuruganathan. Introduction That Mean Shift Algorithm. Informal introduction of the algorithm into McanShift with links for further study: access mode <https://saravananthirumuruganathan.wordpress.com/about/>
3. Documentation library SeikitLcarn. It also contains a huge amount of theoretical information and links for further study: access mode <https://scikit-learn.org/stable/documentation.html>
4. Sebastian Rasehka. Kernel density method. A very clear explanation of what the Parzenetwork window is and what the kernel function is: access mode [http://sebastianraschka.com/Articles/2014\\_kernel\\_density\\_est.html](http://sebastianraschka.com/Articles/2014_kernel_density_est.html)
5. Leo Breiman and Adele Cutler. Random Forests. Description of the Random Forest algorithm from its authors: access mode [https://www.stat.berkeley.edu/~breiman/RandomForests/cc\\_home.htm](https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm)
6. Trevor Hastie Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning. A fundamental introduction to machine learning techniques. The most famous book in this area: access mode <https://web.stanford.edu/~hastie/ElemStatLearn/>
7. Sebastian Rasehka. Python Machine Learning. Practical introduction to machine learning with the presentation: access mode <https://www.amazon.com/Python-Machine-Learning-Sebastian-Raschka/dp/1783555130>
8. Decision tree learning. An article on decision trees in machine learning: access mode [https://en.wikipedia.org/wiki/Decision\\_tree\\_learning](https://en.wikipedia.org/wiki/Decision_tree_learning)
9. Gini Impurity vs Entropy. The question on Stackcxhange with short and clear answers and all the necessary links: access mode <https://datascience.stackexchange.com/questions/10228/when-should-i-use-gini-impurity-as-opposed-to-information-gain>

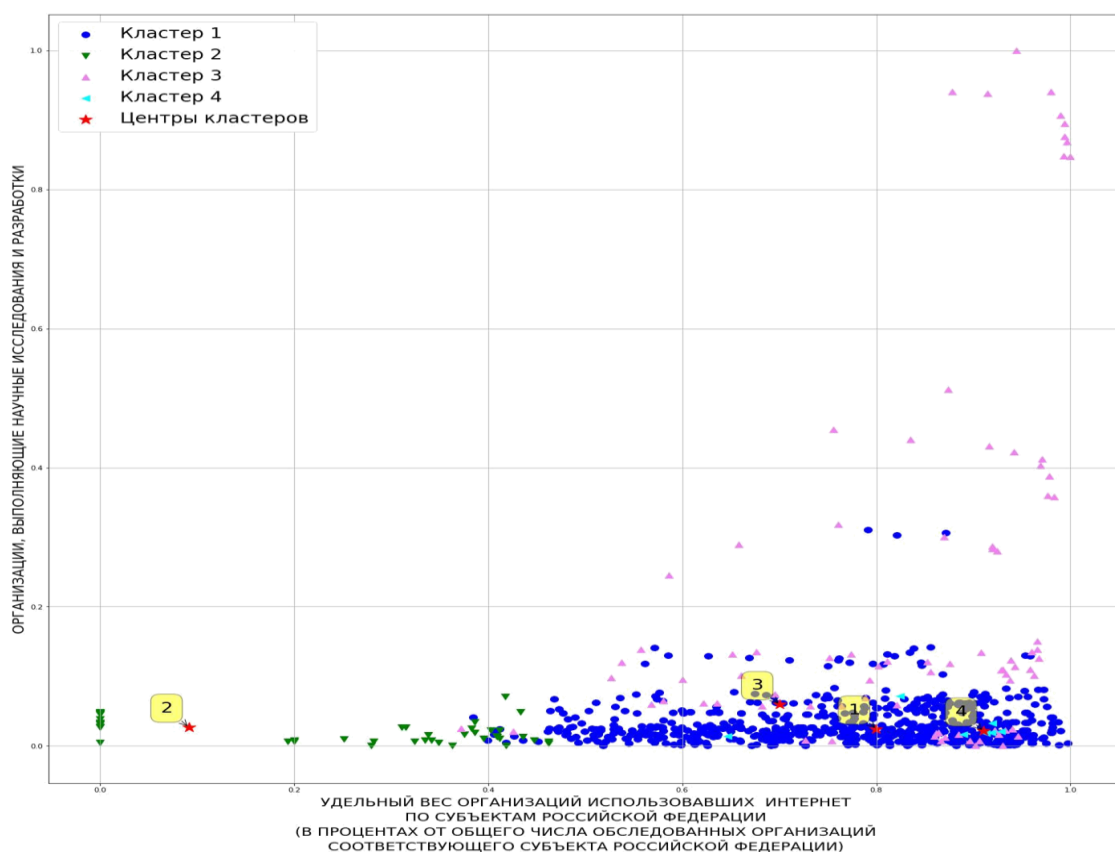
ПРИЛОЖЕНИЕ 1











ПРИЛОЖЕНИЕ 2

Год: 2005



**Кластер: 1**

BEL Белгородская область, VLA Владимирская область, VOR Воронежская область, KAL Калужская область, LIP Липецкая область, RYA Рязанская область, TAM Тамбовская область, YAR Ярославская область, KAR Республика Карелия, КОМ Республика Коми, NAO Ненецкий автономный округ, VOL Вологодская область, KAN Калининградская область, LEN Ленинградская область, MUR Мурманская область, NOV Новгородская область, PSK Псковская область, ADY Республика Адыгея, KRA Краснодарский край, AST Астраханская область, VOG Волгоградская область, ROS Ростовская область, DAG Республика Дагестан,

К.В.Р Кабардино-Балкарская Республика, СТА Ставропольский край, ВАС Республика Башкортостан, MEL Республика Марий Эл, UDM Удмуртская Республика, CHU Чувашская Республика, ORN Оренбургская область, PEN Пензенская область, SAR Саратовская область, КНА Ханты-Мансийский автономный округ, ЯМАО Ямало-Ненецкий автономный округ, CHL Челябинская область, ALT Республика Алтай, КНК Республика Хакасия, АТА Алтайский край, IRK Иркутская область, КЕМ Кемеровская область, NOS Новосибирская область, OMS Омская область, ТОМ Томская область, SAK Республика Саха (Якутия), КАС Камчатский край, PRI Приморский край, КНВ Хабаровский край, МАГ Магаданская область, SAK Сахалинская область, САО Чукотский автономный округ

#### Кластер: 2

БРЯ Брянская область, ИВА Ивановская область, КОС Костромская область, КУР Курская область, СМО Смоленская область, ТВЕ Тверская область, АРК Архангельская область, КАМ Республика Калмыкия, ИНС Республика Ингушетия, К.С.Р Карачаево-Черкесская Республика, СЕО Республика Северная Осетия, ЧНЕ Чеченская Республика, MOR Республика Мордовия, КИР Кировская область, УЛЮ Ульяновская область, КУР Курганская область, ТЮУ Тюменская область, БУР Республика Бурятия, ТЮВ Республика Тыва, ВАИ Забайкальский край, КРС Красноярский край, АМУ Амурская область, .JEW Еврейская автономная область

#### Кластер: 3

МОК Московская область, ОРЕ Орловская область, ТУЛ Тульская область, МОС Москва, СТ.Р Санкт-Петербург, ТАТ Республика Татарстан, PER Пермский край, НИЗ Нижегородская область, САМ Самарская область, СВЕ Свердловская область

#### Кластер: 4

Год: 2006



#### Кластер: 1

ВЕС Белгородская область, БРЯ Брянская область, ВЛА Владимирская область, ВОР Воронежская область, ИВА Ивановская область, КАЛ Калужская область, КОС Костромская область, КУР Курская область, ЛИП Липецкая область, ОРЕ Орловская область, РYA Рязанская область, ТАМ Тамбовская область, ТВЕ Тверская область, ТУЛ Тульская область, ЯАР Ярославская область, КАР Республика Карелия, КОМ Республика Коми, НАО Ненецкий автономный округ, ВОЛ Вологодская область, КАЛ Калининградская область, ЛЕН Ленинградская область, МУР Мурманская область, НОВ Новгородская область, ПСК Псковская область, АДЮ Республика Адыгея, КРА Краснодарский край, АСТ Астраханская область, ВОГ Волгоградская область, РОС Ростовская область, ДАГ Республика Дагестан,



К.В.Р Кабардино-Балкарская Республика, К.С.Р Карачаево-Черкесская Республика, СЕО Республика Северная Осетия, СТА Ставропольский край, ВАС Республика Башкортостан, МЕЛ Республика Марий Эл, MOR Республика Мордовия, UDM Удмуртская Республика, CHU Чувашская Республика, ORN Оренбургская область, PEN Пензенская область, SAR Саратовская область, ULY Ульяновская область, KUR Курганская область, КНА Ханты-Мансийский автономный округ, ЯМАО Ямало-Ненецкий автономный округ, CHL Челябинская область, ALT Республика Алтай, BUR Республика Бурятия, КНК Республика Хакасия, АТА Алтайский край, ВАИ Забайкальский край, KRS Красноярский край, ИРК Иркутская область, КЕМ Кемеровская область, NOS Новосибирская область, OMS Омская область, ТОМ Томская область, SAK Республика Саха (Якутия), КАС Камчатский край, ПРИ Приморский край, КНВ Хабаровский край, АМУ Амурская область, МАГ Магаданская область, САК Сахалинская область, .JEW Еврейская автономная область, САО Чукотский автономный округ

#### Кластер: 2

СМО Смоленская область, АРК Архангельская область, КАМ Республика Калмыкия, ИНГ Республика Ингушетия, СЧЕ Чеченская Республика, КИР Кировская область, ТЮУ Тюменская область, ТЮВ Республика Тыва

#### Кластер: 3

МОК Московская область, МОС Москва, СТ.Р Санкт-Петербург, ТАТ Республика Татарстан, ПЕР Пермский край, НИЗ Нижегородская область, САМ Самарская область, СВЕ Свердловская область

#### Кластер: 4

Год: 2007



#### Кластер: 1

ВЕР Белгородская область, ВБР Брянская область, ВЛА Владимирская область, ВОР Воронежская область, ИВА Ивановская область, КАЛ Калужская область, КОС Костромская область, КУР Курская область, ЛИП Липецкая область, ОРЕ Орловская область, РЯА Рязанская область, СМО Смоленская область, ТАМ Тамбовская область, ТВЕ Тверская область, ТУЛ Тульская область, ЯАР Ярославская область, КАР Республика Карелия, КОМ Республика Коми, НАО Ненецкий автономный округ, ВОЛ Вологодская область, КАК Калининградская область, ЛЕН Ленинградская область, МУР Мурманская область, НОВ Новгородская область, ПСК Псковская область, АДЫ Республика Адыгея, КАМ Республика

Калмыкия, KRA Краснодарский край, AST Астраханская область, VOG Волгоградская область, ROS Ростовская область, DAG Республика Дагестан, К.В.Р Кабардино-Балкарская Республика, К.С.Р Карачаево-Черкесская Республика, СЕО Республика Северная Осетия, STA Ставропольский край, В AS Республика Башкортостан, MEL Республика Марий Эл, UDM Удмуртская Республика, CHU Чувашская Республика, KIR Кировская область, ORN Оренбургская область, PEN Пензенская область, SAR Саратовская область, ULY Ульяновская область, KUR Курганская область, КНА Ханты-Мансийский автономный округ, YAM Ямало-Ненецкий автономный округ, CHL Челябинская область, ALT Республика Алтай, BUR Республика Бурятия, КНК Республика Хакасия, АТА Алтайский край, ВАИ Забайкальский край, KRS Красноярский край, IRK Иркутская область, КЕМ Кемеровская область, NOS Новосибирская область,

OMS Омская область, ТОМ Томская область, SAK Республика Саха (Якутия), КАС Камчатский край, PRI Приморский край, КНВ Хабаровский край, АМУ Амурская область, MAG Магаданская область, SAK Сахалинская область, JEW Еврейская автономная область, САО Чукотский автономный округ **Кластер: 2**

ARK Архангельская область, ING Республика Ингушетия, СЧЕ Чеченская Республика, ТЮУ Тюменская область, ТЮВ Республика Тыва

#### Кластер: 3

МОК Московская область, МОС Москва, ST.P Санкт-Петербург, ТАТ Республика Татарстан, PER Пермский край, NIZ Нижегородская область, САМ Самарская область, SVE Свердловская область

#### Кластер: 4

MOR Республика Мордовия

Год: 2008



#### Кластер: 1

BEL Белгородская область, BRY Брянская область, VLA Владимирская область, VOR Воронежская область, IVA Ивановская область, KAL Калужская область, KOS Костромская область, KUR Курская область, LIP Липецкая область, МОК Московская область, ORE Орловская область, RYA Рязанская область, SMO Смоленская область, ТАМ Тамбовская область, ТУЕ Тверская область, TUL Тульская область, YAR Ярославская область, KAR Республика Карелия, КОМ Республика Коми, NAO Ненецкий автономный округ, VOL Вологодская область, KAN Калининградская область, LEN Ленинградская область, MUR Мурманская область, NOV Новгородская область, PSK Псковская область, ADY Республика Адыгея, КАМ

Республика Калмыкия, KRA Краснодарский край, AST Астраханская область, VOG Волгоградская область, ROS Ростовская область, DAG Республика Дагестан, ING Республика Ингушетия, К.В.Р Кабардино-Балкарская Республика, К.С.Р Карачаево-Черкесская Республика, СЕО Республика Северная Осетия, СТА Ставропольский край, BAS Республика Башкортостан, MEL Республика Марий Эл, MOR Республика Мордовия, UDM Удмуртская Республика,

' CHU Чувашская Республика, KIR Кировская область, ORN Оренбургская область, PEN Пензенская область, SAR Саратовская область, ULY Ульяновская область, KUR Курганская область, КНА Ханты-Мансийский автономный округ, ЯМ Ямало-Ненецкий автономный округ, CHL Челябинская область, ALT Республика Алтай, BUR Республика Бурятия, КНК Республика Хакасия, АТА Алтайский край, ВАИ Забайкальский край, KRS Красноярский край, IRK Иркутская область, КЕМ Кемеровская область, NOS Новосибирская область, OMS Омская область, ТОМ Томская область, SAK Республика Саха (Якутия), KAC Камчатский край, PRI Приморский край, КНВ Хабаровский край, AMU Амурская область, SAK Сахалинская область, .JEW Еврейская автономная область, С АО Чукотский автономный округ

#### Кластер: 2

ARK Архангельская область, СНЕ Чеченская Республика, ТЮУ Тюменская область, ТУВ Республика Тыва

#### Кластер: 3

MOS Москва, ST.P Санкт-Петербург, ТАТ Республика Татарстан, PER Пермский край, NIZ Нижегородская область, SAM Самарская область, SVE Свердловская область, MAG Магаданская область

#### Кластер: 4

Год: 2009



#### Кластер: 1

BEL Белгородская область, BRY Брянская область, VLA Владимирская область, VOR Воронежская область, IVA Ивановская область, KAL Калужская область, KOS Костромская область, KUR Курская область, LIP Липецкая область, МОК Московская область, ORE Орловская область, RYA Рязанская область, SMO Смоленская область, ТАМ Тамбовская область, ТУЕ Тверская область, TUL Тульская область, YAR Ярославская область, KAR Республика Карелия, КОМ Республика Коми, NAO Ненецкий автономный округ, VOL Вологодская область, KAN Калининградская область, LEN Ленинградская область, MUR Мурманская область, NOV Новгородская область, PSK Псковская область, АДЫ Республика Адыгея, КАМ Республика Калмыкия, KRA Краснодарский край, AST Астраханская область, VOG Волго-

Градская область, ROS Ростовская область, DAG Республика Дагестан, ING Республика Ингушетия, К.В.Р Кабардино-Балкарская Республика, К.С.Р Карачаево-Черкесская Республика, СЕО Республика Северная Осетия, СЧЕ Чеченская Республика, СТА Ставропольский край, В AS Республика Башкортостан, MEL Республика Марий Эл, MOR Республика Мордовия, UDM Удмуртская Республика, CHU Чувашская Республика, KIR Кировская область, ORN Оренбургская область, PEN Пензенская область, SAM Самарская область, SAR Саратовская область, ULY Ульяновская область, KUR Курганская область, КНА Ханты-Мансийский автономный округ, ЯМАО Ямало-Ненецкий автономный округ, CHL Челябинская область, ALT Республика Алтай, BUR Республика Бурятия, ТУВ Республика Тыва, КНК Республика Хакасия, АТА Алтайский край, ВАИ Забайкальский край, KRS Красноярский край, ИРК Иркутская область, КЕМ Кемеровская область, NOS Новосибирская область, OMS Омская область, ТОМ Томская область, SAK Республика Саха (Якутия), KAC Камчатский край, PRI Приморский край, КНВ Хабаровский край, АМУ Амурская область, SAK Сахалинская область, .JEW Еврейская автономная область, САО Чукотский автономный округ

**Кластер: 2**

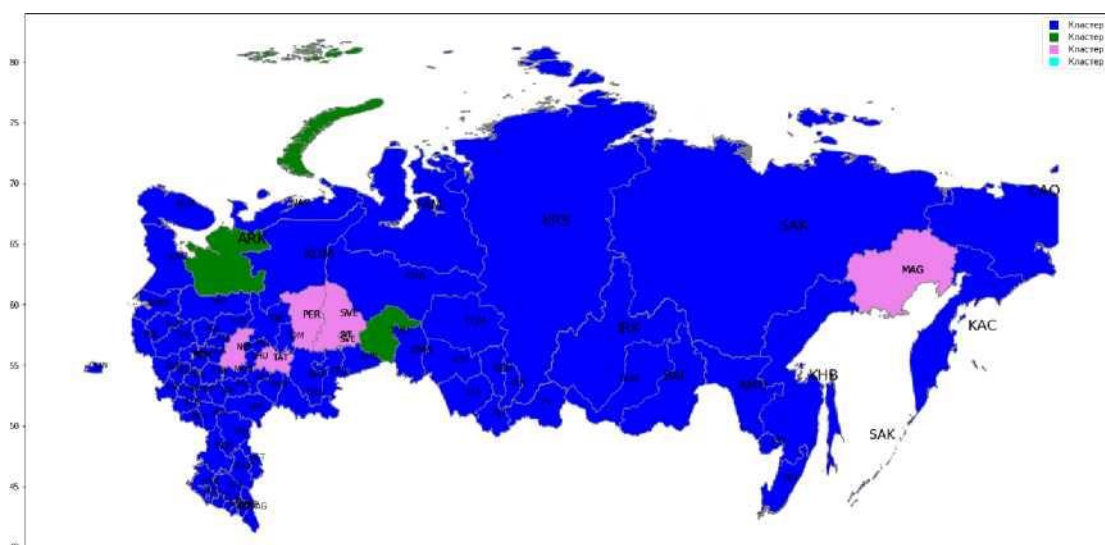
ARK Архангельская область, ТЮТ Тюменская область

**Кластер: 3**

MOS Москва, ST.P Санкт-Петербург, ТАТ Республика Татарстан, PER Пермский край, NIZ Нижегородская область, SVE Свердловская область, МАГ Магаданская область

**Кластер: 4**

Год: 2010

**Кластер: 1**

BEL Белгородская область, ВРБ Брянская область, ВЛА Владимирская область, ВОР Воронежская область, ИВА Ивановская область, КАЛ Калужская область, КОС Костромская область, КУР Курская область, ЛИП Липецкая область, МОК Московская область, ОРЛ Орловская область, РЯА Рязанская область, СМО Смоленская область, ТАМ Тамбовская область, ТУЕ Тверская область, ТУЛ Тульская область, ЯРО Ярославская область, КАР Республика Карелия, КОМ Республика Коми, НАО Ненецкий автономный округ, ВОЛ Вологодская область, КАК Калининградская область, ЛЕН Ленинградская область, МУР Мурманская область, НОВ Новгородская область, ПСК Псковская область, АДЫ Республика Адыгея, КАМ Республика Калмыкия, КРА Краснодарский край, АСТ Астраханская область, ВОГ Волгоградская область, ROS Ростовская область, DAG Республика Дагестан, ING Республика

Ингушетия, К.В.Р Кабардино-Балкарская Республика, К.С.Р Карачаево-Черкесская Республика, СЕО Республика Северная Осетия, СЧЕ Чеченская Республика, СТА Ставропольский край, В АС Республика Башкортостан, МЕЛ Республика Марий Эл, MOR Республика Мордовия, UDM Удмуртская Республика, СЧУ Чувашская Республика, КИР Кировская область, ОРН Оренбургская область, PEN Пензенская область, САМ Самарская область, САР Саратовская область, УЛЮ Ульяновская область, КУР Курганская область, КНА Ханты-Мансийский автономный округ, ЯАМ Ямало-Ненецкий автономный округ, СЧЛ Челябинская область, АЛТ Республика Алтай, BUR Республика Бурятия, ТЫВ Республика Тыва, КНК Республика Хакасия, АТА Алтайский край, ВАИ Забайкальский край, КРС Красноярский край, ИРК Иркутская область, КЕМ Кемеровская область, НОС Новосибирская область, ОМС Омская область, ТОМ Томская область, САК Республика Саха (Якутия), КАС Камчатский край, ПРИ Приморский край, КНВ Хабаровский край, АМУ Амурская область, САК Сахалинская область, .JEW Еврейская автономная область, С АО Чукотский автономный округ

**Кластер: 2**

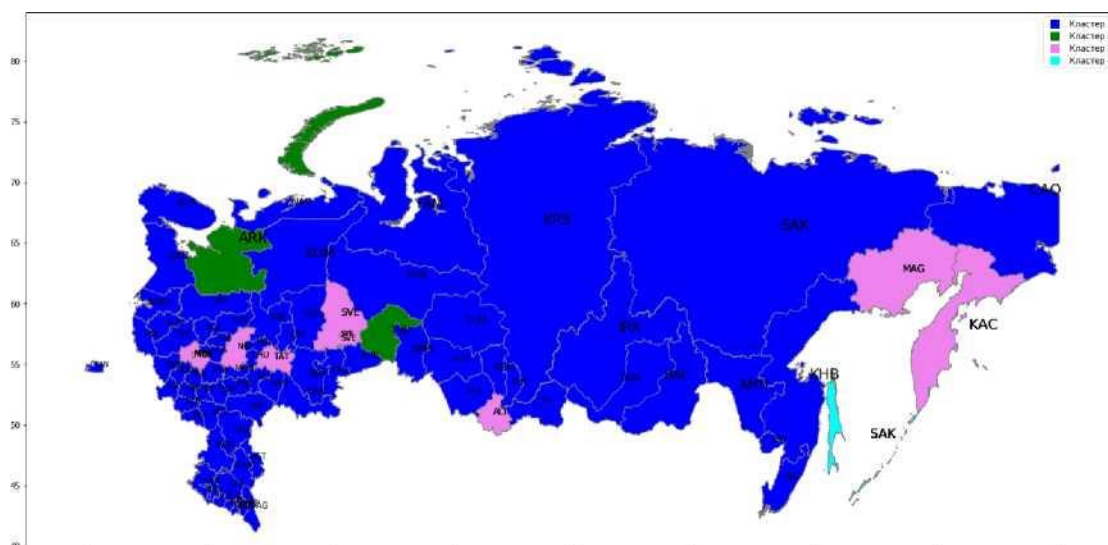
АРК Архангельская область, ТЮУ Тюменская область

**Кластер: 3**

МОС Москва, СТ.Р Санкт-Петербург, ТАТ Республика Татарстан, PER Пермский край, НИЗ Нижегородская область, SVE Свердловская область, МАГ Магаданская область

**Кластер: 4**

Год: 2011

**Кластер: 1**

BEL Белгородская область, ВРУ Брянская область, VLA Владимирская область, VOR Воронежская область, IVA Ивановская область, KAL Калужская область, KOS Костромская область, КУР Курская область, ЛIP Липецкая область, ORE Орловская область, РYA Рязанская область, SMO Смоленская область, ТАМ Тамбовская область, TVE Тверская область, TUL Тульская область, ЯАР Ярославская область, KAR Республика Карелия, КОМ Республика Коми, NAO Ненецкий автономный округ, VOL Вологодская область, К AN Калининградская область, LEN Ленинградская область, MUR Мурманская область, NOV Новгородская область, PSK Псковская область, АДY Республика Адыгея, КАМ Республика Калмыкия, KRA Краснодарский край, АСТ Астраханская область, VOG Волгоградская область, ROS Ростовская область, DAG Республика Дагестан, ING Республика Ингушетия, К.В.Р Кабардино-Балкарская Республика, К.С.Р Карачаево-Черкесская Республика, СЕО Республика Северная Осетия, СЧЕ Чеченская Республика, СТА Ставропольский край, ВАС Республика Башкортостан, МЕЛ Республика Марий Эл, MOR Республика Мордовия, UDM Удмуртская

Республика, CHU Чувашская Республика, PER Пермский край, KIR Кировская область, ORN Оренбургская область, PEN Пензенская область, SAM Самарская область, SAR Саратовская область, ULY Ульяновская область, KUR Курганская область, КНА Ханты-Мансийский автономный округ, ЯАМ Ямало-Ненецкий автономный округ, CHL Челябинская область, BUR Республика Бурятия, TYV Республика Тыва, КНК Республика Хакасия, АТА Алтайский край, ВАИ Забайкальский край, KRS Красноярский край, IRK Иркутская область, КЕМ Кемеровская область, NOS Новосибирская область, OMS Омская область, ТОМ Томская область, SAK Республика Саха (Якутия), PRI Приморский край, КНВ Хабаровский край, АМУ Амурская область, JEW Еврейская автономная область, САО Чукотский автономный округ

#### Кластер: 2

ARK Архангельская область, TYU Тюменская область

#### Кластер: 3

МОК Московская область, MOS Москва, ST.P Санкт-Петербург, ТАТ Республика Татарстан, NIZ Нижегородская область, SVE Свердловская область, АЛТ Республика Алтай, КАС Камчатский край, МАГ Магаданская область

#### Кластер: 4

SAK Сахалинская область

Год: 2012



#### Кластер: 1

BEL Белгородская область, BRY Брянская область, VLA Владимирская область, VOR Воронежская область, IVA Ивановская область, KAL Калужская область, KOS Костромская область, KUR Курская область, LIP Липецкая область, ORE Орловская область, RYA Рязанская область, SMO Смоленская область, TAM Тамбовская область, TVE Тверская область, TUL Тульская область, YAR Ярославская область, KAR Республика Карелия, КОМ Республика Коми, NAO Ненецкий автономный округ, VOL Вологодская область, KAN Калининградская область, LEN Ленинградская область, MUR Мурманская область, NOV Новгородская область, PSK Псковская область, АДY Республика Адыгея, КАМ Республика Калмыкия, KRA Краснодарский край, AST Астраханская область, VOG Волгоградская область, ROS Ростовская область, DAG Республика Дагестан, ING Республика Ингушетия, К.В.Р Кабардино-Балкарская Республика, К.С.Р Карачаево-Черкесская Республика, СЕО Республика Северная Осетия, СНЕ Чеченская Республика, СТА Ставропольский край, ВАС Республика Башкортостан, MEL Республика Марий Эл, MOR Республика Мордовия, UDM Удмуртская Республика, PER Пермский край, KIR Кировская область, ORN Оренбургская область, PEN

Пензенская область, SAR Саратовская область, ULY Ульяновская область, KUR Курганская область, ХНА Ханты-Мансийский автономный округ, ЯМ Ямало-Ненецкий автономный округ, CHL Челябинская область, ALT Республика Алтай, BUR Республика Бурятия, ТУВ Республика Тыва, КНК Республика Хакасия, АТА Алтайский край, ВАИ Забайкальский край, KRS Красноярский край, IRK Иркутская область, КЕМ Кемеровская область, NOS Новосибирская область, OMS Омская область, ТОМ Томская область, SAK Республика Саха (Якутия), PRI Приморский край, КНВ Хабаровский край, АМУ Амурская область, .JEW Еврейская автономная область, САО Чукотский автономный округ

**Кластер: 2**

ARK Архангельская область, ТУУ Тюменская область

**Кластер: 3**

МОК Московская область, МОС Москва, СТ.Р Санкт-Петербург, ТАТ Республика Татарстан, CHU Чувашская Республика, NIZ Нижегородская область, SVE Свердловская область, КАС Камчатский край, МАГ Магаданская область

**Кластер: 4**

SAM Самарская область, SAK Сахалинская область

Год: 2013

**Кластер: 1**

BEL Белгородская область, BRY Брянская область, VLA Владимирская область, VOR Воронежская область, IVA Ивановская область, KAL Калужская область, KOS Костромская область, KUR Курская область, LIP Липецкая область, ORE Орловская область, RYA Рязанская область, SMO Смоленская область, ТАМ Тамбовская область, TVE Тверская область, TUL Тульская область, YAR Ярославская область, KAR Республика Карелия, КОМ Республика Коми, NAO Ненецкий автономный округ, VOL Вологодская область, KAN Калининградская область, LEN Ленинградская область, MUR Мурманская область, NOV Новгородская область, PSK Псковская область, АДЫ Республика Адыгея, КАМ Республика Калмыкия, KRA Краснодарский край, AST Астраханская область, VOG Волгоградская область, ROS Ростовская область, DAG Республика Дагестан, ING Республика Ингушетия, К.В.Р Кабардино-Балкарская Республика, К.С.Р Карачаево-Черкесская Республика, СЕО Республика Северная Осетия, СНЕ Чеченская Республика, СТА Ставропольский край, BAS Республика Башкортостан, MEL Республика Марий Эл, UDM Удмуртская Республика, PER Пермский край, KIR Кировская область, ORN Оренбургская область, PEN Пензенская область, SAM Самарская область, SAR Саратовская область, ULY Ульяновская область, KUR Курганская область, SVE Свердловская область, ТУУ Тюменская область, ХНА Ханты-Мансийский

автономный округ, YAM Ямало-Ненецкий автономный округ, CHL Челябинская область, ALT Республика Алтай, BUR Республика Бурятия, TYV Республика Тыва, KHK Республика Хакасия, ATA Алтайский край, BAI Забайкальский край, KRS Красноярский край, IRK Иркутская область, KEM Кемеровская область, NOS Новосибирская область, OMS Омская область, TOM Томская область, SAK Республика Саха (Якутия), KAC Камчатский край, PRI Приморский край, KHB Хабаровский край, AMU Амурская область, JEW Еврейская автономная область

**Кластер: 2****Кластер: 3**

МОК Московская область, MOS Москва, ST.P Санкт-Петербург, MOR Республика Мордовия, TAT Республика Татарстан, CHU Чувашская Республика, NIZ Нижегородская область, MAG Магаданская область, CAO Чукотский автономный округ

**Кластер: 4**

ARK Архангельская область, SAK Сахалинская область

Год: 2014

**Кластер: 1**

BEL Белгородская область, BRY Брянская область, VLA Владимирская область, VOR Воронежская область, IVA Ивановская область, KAP Калужская область, KOS Костромская область, KUR Курская область, ORE Орловская область, RYA Рязанская область, SMO Смоленская область, TAM Тамбовская область, TVE Тверская область, TUL Тульская область, YAR Ярославская область, KAR Республика Карелия, КОМ Республика Коми, ARK Архангельская область, NAO Ненецкий автономный округ, VOL Вологодская область, KAN Калининградская область, LEN Ленинградская область, MUR Мурманская область, NOV Новгородская область, PSK Псковская область, ADY Республика Адыгея, КАМ Республика Калмыкия, KRA Краснодарский край, AST Астраханская область, VOG Волгоградская область, ROS Ростовская область, DAG Республика Дагестан, ING Республика Ингушетия, К.В.Р Кабардино-Балкарская Республика, К.С.Р Карачаево-Черкесская Республика, СЕО Республика Северная Осетия, СНЕ Чеченская Республика, STA Ставропольский край, BAS Республика Башкортостан, MEL Республика Марий Эл, UDM Удмуртская Республика, PER Пермский край, KIR Кировская область, ORN Оренбургская область, PEN Пензенская область, SAM Самарская область, SAR Саратовская область, ULY Ульяновская область, KUR Курганская область, SVE Свердловская область, TYU Тюменская область, КНА Ханты-Мансийский автономный округ, YAM Ямало-Ненецкий автономный округ, CHL Челябинская область, ALT Республика Алтай, BUR Республика Бурятия, TYV Республика



Тыва, КНК Республика Хакасия, АТА Алтайский край, ВАІ Забайкальский край, KRS Красноярский край, ІRK Иркутская область, КЕМ Кемеровская область, NOS Новосибирская область, OMS Омская область, ТОМ Томская область, SAK Республика Саха (Якутия), KAC Камчатский край, PRI Приморский край, КНВ Хабаровский край, АМУ Амурская область, МАG Магаданская область, JEW Еврейская автономная область

**Кластер: 2**

**Кластер: 3**

LIP Липецкая область, МОК Московская область, MOS Москва, ST.P Санкт-Петербург, MOR Республика Мордовия, TAT Республика Татарстан, CHU Чувашская Республика, NIZ Нижегородская область, С АО Чукотский автономный округ

**Кластер: 4**

SAK Сахалинская область