Уфимский государственный авиационный технический университет

«Интеллектуальный анализ основных показателей здоровья матери и ребенка на территории Российской Федерации за 2018 год»

Рафикова А.А., магистр группы БТС-207м

Российская Федерация

**Содержание**

[Аннотация 3](#_Toc55697167)

[Практическая часть исследования 5](#_Toc55697168)

[Вывод 15](#_Toc55697169)

[Список литературы 16](#_Toc55697170)

# Аннотация

Вопросы охраны материнского, перинатального здоровья и здоровья новорожденных касаются общества в целом и каждой страны в отдельности. В Российской Федерации данный вопрос стоит остро, учитывая сокращение рождаемости. Достижение ЦТР 5 (улучшение охраны материнства) является крайне важным для достижения ЦТР 4 (сокращение детской смертности).

Под здоровьем матери подразумевается здоровье женщины во время вынашивания ребенка, родов и в послеродовой период. Перинатальное здоровье – это состояние здоровья ребенка, которое отсчитывается с полных 22 недель беременности матери, и до того момента, когда после рождения малыша прошло полных 7 дней. Здоровье новорожденных – это здоровье младенца в первый месяц жизни. Правильный подход к началу жизни плода в перинатальный период, несомненно, влияет на здоровье в младенчестве и последующих периодах жизни.

В данной статье выполняется интеллектуальный анализ данных об основных показателях здоровья матери и ребенка на территории Российской Федерации за 2017 год методами компонентного, кластерного анализа, построения деревьев решений, а также нейросетевыми методами.

Цель анализа состоит в выявлении закономерностей вида классификации при помощи метода главных компонент и кластеризации с использованием карт Кохонена. Основная задача \_ выявить различия между регионами РФ по показателям здоровья матери и ребенка на территории Российской Федерации.

Анализ будет проведен по следующим признакам:

1. Количество гинекологических коек на 10000 женского населения (*Gynaecology* *beds*), %;
2. Количество патологий беременности (*Disease during pregnancy*);
3. Доля беременных, поступивших под наблюдение консультации до 12 недель беременности (*Prenatal observation*), %
4. Процент родов в надлежащих условиях (*Delivery in the maternity home*), %;
5. Количество врачей-педиатров (*Pediatrician*);
6. Количество новорожденных детей (*New-borns*);
7. Количество мертворожденных детей (*Stillbirth*);
8. Круглосуточные стационарные койки для детей (*In-patient admissions*).

# Практическая часть исследования

Программные продукты

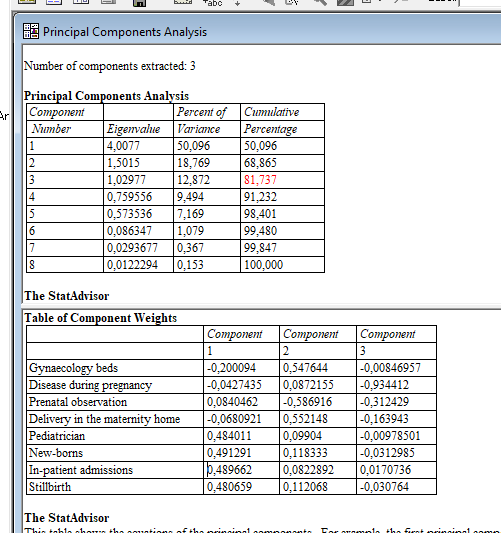
Для выполнения исследования были использованы программные пакеты: *Deductor Studio Academic*, *STATGRAPHIC Centurion*.

Анализ методом главных компонент

Метод главных компонент (ГК) осуществляет переход к новой системе координат в исходном пространстве признаков, являющейся системой ортонормированных линейных комбинаций. Метод ГК заключается в нахождении набора из p ортогональных векторов в n-мерном исходном пространстве данных; поскольку при этом p<<n, то это дает возможность перейти к сокращенному признаковому пространству. Эффективность метода состоит в минимальном искажении геометрической структуры точек (объектов) при их проектировании в пространство меньшей размерности.

Введем исходные данные в электронную таблицу *STATGRAPHICS* (85\*8). Исходная сводка анализа метода ГК представлена в таблице 1.

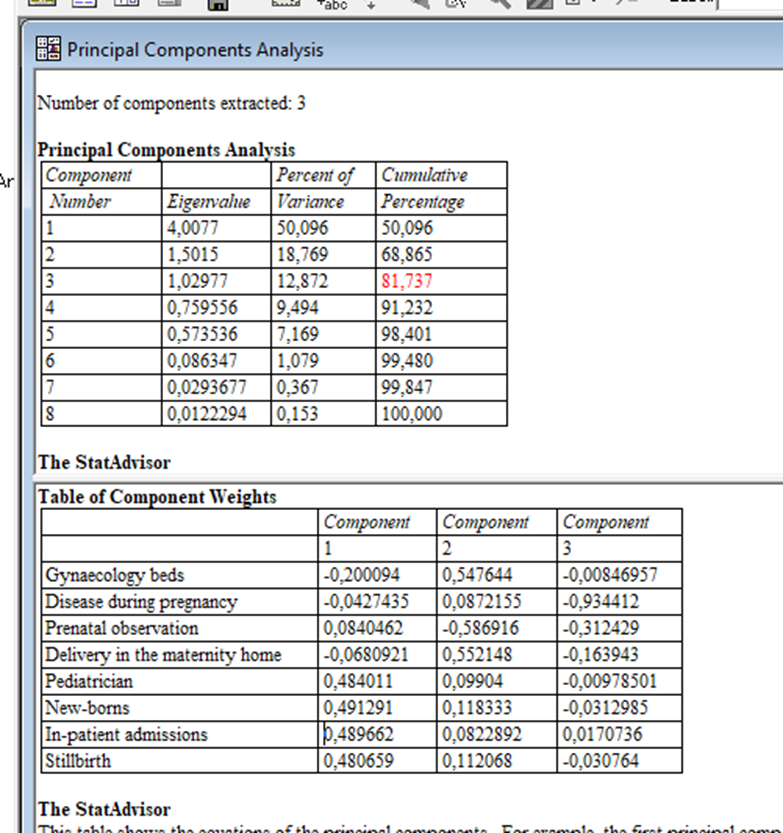
Таблица 1 \_ Исходная сводка анализа метода ГК



Приведенные цифры говорят о том, что уже первые две главные компоненты описывают 68,865% дисперсии исходных данных. Третья главная компонента добавляет еще 12,872% дисперсии, так что в сумме получается 81,737% дисперсии, что достаточно для анализа пространственного распределения объектов.

В следующей таблице представлены веса признаков в главных компонентах:

Таблица 2 \_ Веса признаков в главных компонентах



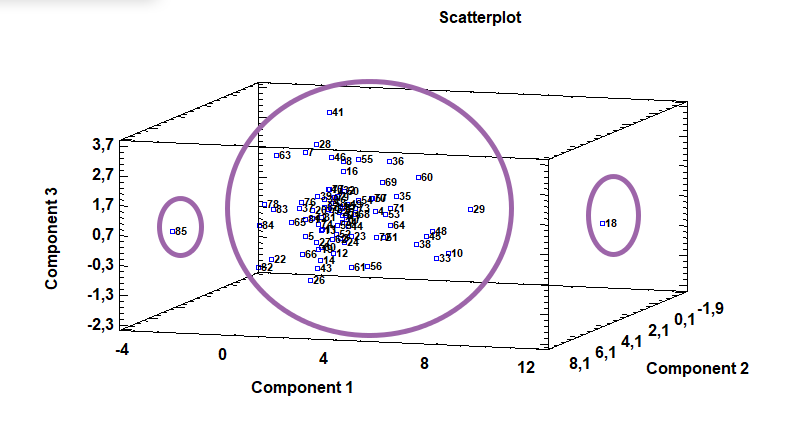
Как следует из таблицы (берем во внимание наибольшие числа), первая главная компонента в наибольшей степени зависит от следующих признаков: количество врачей-педиатров, количество новорожденных детей, количество мертворожденных детей, круглосуточные стационарные койки для детей. Название 1-й главной компоненты «Жизнеспособность детей в зависимости от больничных условий».

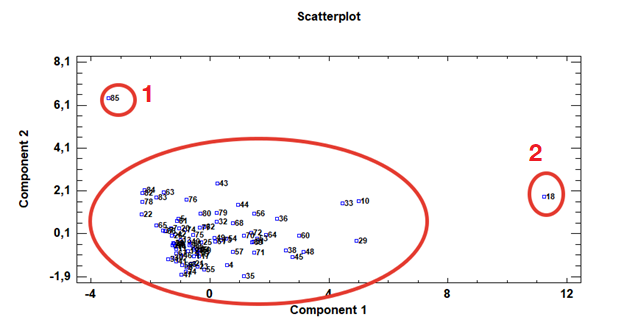
2-я главная компонента «Наблюдение за беременными женщинами» в наибольшей степени зависит от следующих признаков: количество гинекологических коек на 10000 женского населения; доля беременных, поступивших под наблюдение консультации до 12 недель беременности; процент родов в надлежащих условиях.

3-я главная компонента «Количество беременных женщин с патологиями» зависит от признаков: доля беременных, поступивших под наблюдение консультации до 12 недель беременности; количество патологий беременности.

Выделение кластеров на диаграмме рассеивания

Диаграмма рассеивания всего множества объектов на плоскости выделенных трех главных компонент представлена на рисунке 1:





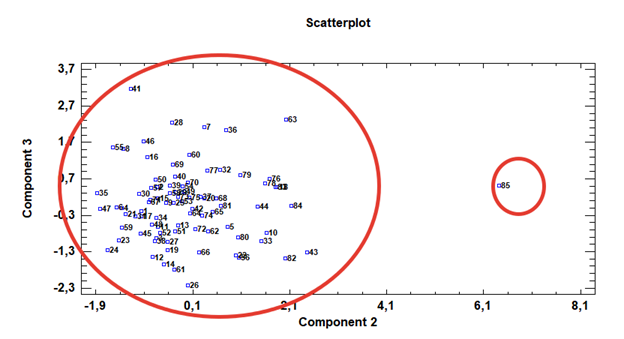


Рисунок 1 \_ Проекция исследуемых объектов на пространство трех ГК

На диаграммах мы можем увидеть четкое разделение всей статистической выборки на три четких класса. Основное количество регионов находится в первом классе.

Формирование правил в пространстве главных компонент:

Таблица 3 \_ Правила

|  |  |  |  |
| --- | --- | --- | --- |
| Класс | Значение компоненты | | |
| Компонента 1 | Компонента 2 | Компонента 3 |
| 1 | низкое | высокое | среднее |
| 2 | высокое | низкое | среднее |
| 3 | высокое | среднее | среднее |

Перечислим состав объектов в выделенных классах.

В первый класс входят номера объектов: 1-17, 19-84.

Во второй класс входит объект: 18.

В третий класс входит объект 85.

Характеристика классов относительно признаков представлена в таблице 4.

Таблица 4

Характеристика классов в пространстве исходных признаков

|  |  |  |  |
| --- | --- | --- | --- |
| Значение признаков | Класс | | |
| 1 | 2 | 3 |
| Количество гинекологических коек на 10000 женского населения | высокое | низкое | низкое |
| Количество патологий беременности | среднее | среднее | высокое |
| Доля беременных, поступивших под наблюдение консультации до 12 недель беременности | низкое | высокое | среднее |
| Процент родов в надлежащих условиях | высокое | низкое | низкое |
| Количество врачей-педиатров | среднее | среднее | низкое |
| Количество новорожденных детей | среднее | низкое | низкое |
| Количество мертворожденных детей | среднее | низкое | среднее |
| Круглосуточные стационарные койки для детей | среднее | среднее | низкое |

Извлеченные правила:

Если Количество гинекологических коек на 10000 женского населения=высокое; Количество патологий беременности=среднее; Доля беременных, поступивших под наблюдение консультации до 12 недель беременности=низкое; Процент родов в надлежащих условиях=высокое; Количество врачей-педиатров=среднее; Количество новорожденных детей= среднее; Количество мертворожденных детей=среднее; Круглосуточные стационарные койки для детей = среднее, то Класс 1.

Если Количество гинекологических коек на 10000 женского населения=низкое; Количество патологий беременности=среднее; Доля беременных, поступивших под наблюдение консультации до 12 недель беременности=высокое; Процент родов в надлежащих условиях= низкое; Количество врачей-педиатров= среднее; Количество новорожденных детей=низкое; Количество мертворожденных детей=среднее; Круглосуточные стационарные койки для детей =среднее, то Класс 2.

Если Количество гинекологических коек на 10000 женского населения=низкое; Количество патологий беременности=высокое; Доля беременных, поступивших под наблюдение консультации до 12 недель беременности=среднее; Процент родов в надлежащих условиях=низкое; Количество врачей-педиатров=низкое; Количество новорожденных детей=низкое; Количество мертворожденных детей=среднее; Круглосуточные стационарные койки для детей=низкое, то Класс 3.

К первому классу относятся 83 региона Российской Федерации.

Ко второму классу относится Ярославская область.

К третьему классу относится Еврейская автономная область.

Кластерный анализ с применением карт Кохонена

Для решения задачи кластеризации используется метод самоорганизующихся карт Кохонена с реализацией на основе нейросетевого пакета *Deductor Studio Academic*.

Самоорганизующиеся карты Кохонена представляют собой прямонаправленные нейронные сети, которые используют алгоритм обучения без учителя, и с помощью процесса самоорганизации формируют выходные ячейки в топологическую карту, имеющую двухмерную структуру.

Карты Кохонена дают более развернутую картину ситуации.

*Первый этап.* Настройка параметров обучений карты Кохонена

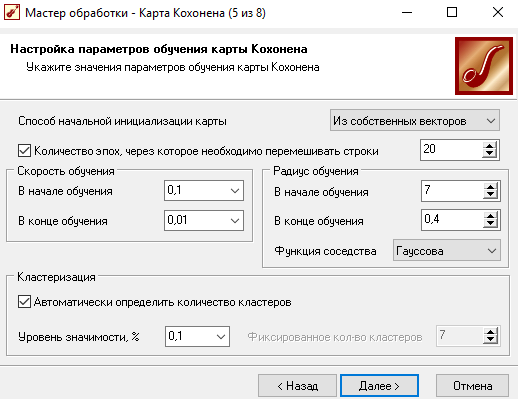


Рисунок 2 \_ Мастер обработки

*Второй этап.* Для построения карт необходимо открыть окно с картами по обучающей выборке и выбрать нужные карты. На картах Кохонена выделяются шесть кластеров. В каждом из этих кластеров сосредотачиваются определенные регионы Российской Федерации. Каждая отдельная карта дает нам представление о том или ином признаке, по которому рассматривается статистическая информация. Например, на карте «Патологии беременности» зеленым цвет преимущественно присутствует на большинстве кластеров и характеризует средний показатель патологий.

Низкий уровень патологий наблюдается в 4 кластере. В нем расположены регионы, указанные на рисунке 5.

Результаты построения карт представлены на рисунке 3.

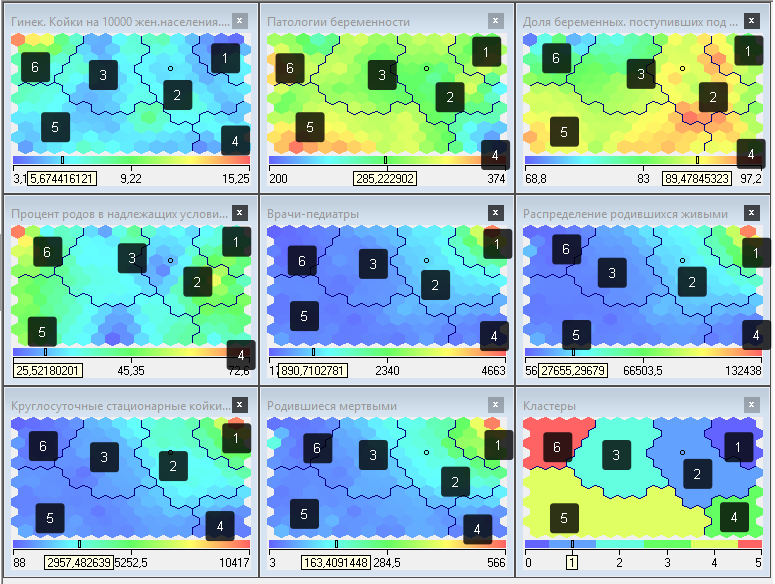


Рисунок 3 \_ Результаты построения карт Кохонена

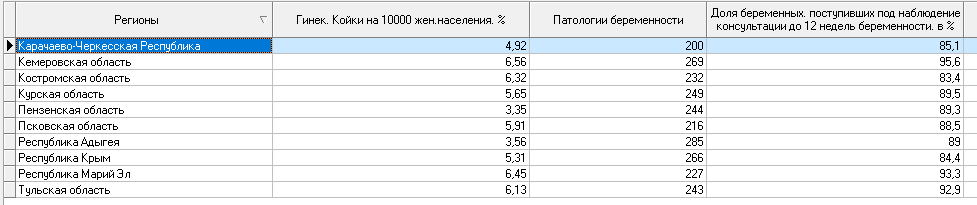
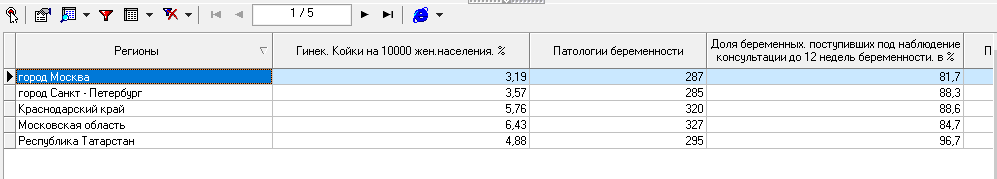
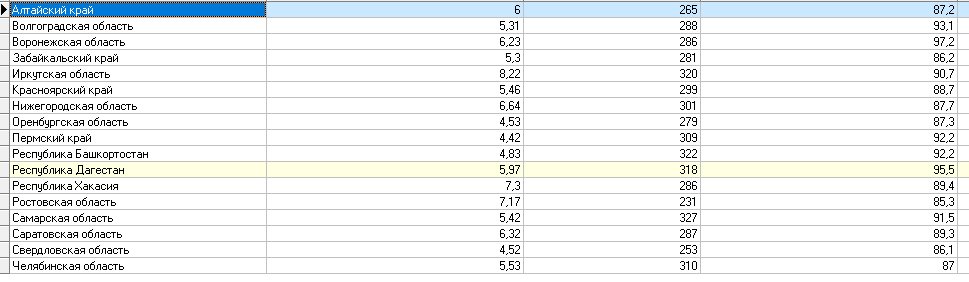


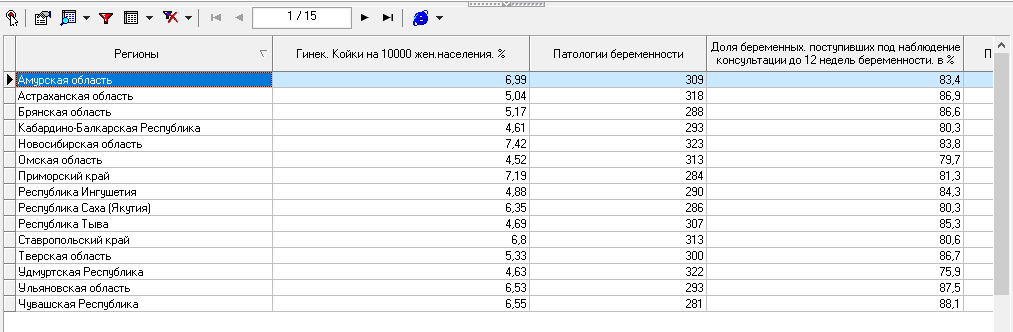
Рисунок 4 \_ Регионы, включенные в 4-й кластер

Регионы, включенные в 1-й кластер:

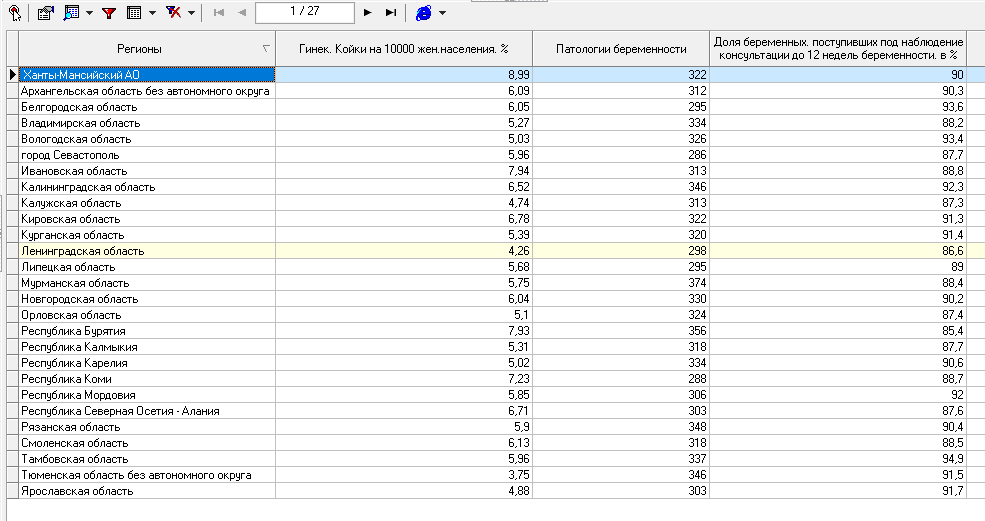


Регионы, включенные во 2-й кластер:

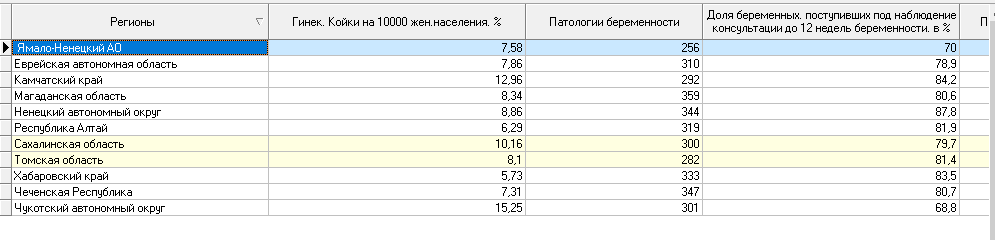


Регионы, включенные в 3-й кластер: 

Регионы, включенные в 4-й кластер:



Регионы, включенные в 5-й кластер:



Таким образом, сформируем таблицу с правилами:

Таблица 5 \_ Правила для карт Кохонена

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Кластеры  Признаки | 1 | 2 | 3 | 4 | 5 | 6 |
| Количество гинекологических коек на 10000 женского населения | низкое | низкое | низкое | низкое | низкое | среднее |
| Количество патологий беременности | среднее | среднее | среднее | низкое | среднее | среднее |
| Доля беременных, поступивших под наблюдение консультации до 12 недель беременности | среднее | высокое | среднее | низкое | среднее | среднее |
| Процент родов в надлежащих условиях | среднее | среднее | низкое | низкое | низкое | среднее |
| Количество врачей-педиатров | высокое | среднее | низкое | низкое | низкое | низкое |
| Количество новорожденных детей | высокое | среднее | низкое | низкое | низкое | низкое |
| Количество мертворожденных детей | высокое | среднее | низкое | низкое | низкое | низкое |
| Круглосуточные стационарные койки для детей | среднее | низкое | низкое | низкое | низкое | низкое |

Полученные данные могут быть использованы в качестве основы для медицинских исследований. По информации, выясненной в ходе работы с картами Кохонена, можно сделать некоторые выводы.

Например, самые низкие показатели у регионов, включенных в 4-й кластер. В данных регионах (Республики Марий Эл, Адыгея, Крым, Карачаево-Черкесская Республика, а также, Тульская, Курская, Кемеровская, Псковская Костромская, и Пензенская области) низкое значение патологий беременности и мало мертворожденных детей, что замечательно, однако, данные регионы плохо обеспечены необходимыми условиями для наблюдения беременных женщин и младенцев. Хорошие показатели здоровья матерей и младенцев могут быть обусловлены южным климатом большинства регионов. Отсутствие условий в некоторых регионах может быть объяснено отдаленностью регионов от центра Российской Федерации.

Также, рассмотрим ситуацию в регионах, включенных в первый кластер (Москва, Санкт-Петербург, Краснодарский край, Республика Татарстан, Московская область). В данных регионах имеется достаточное количество стационарных коек для детей, врачей-педиатров. Высокая рождаемость и смертность младенцев может обуславливаться рассеиванием данных, ведь упомянутые регионы достаточно крупные по площади.

# Вывод

С целью выявления закономерностей вида классификации в рамках исследовательской работы было выполнено формирование знаний с помощью технологий интеллектуального анализа данных путем применения метода главных компонент и нейросетевых методов построения карт Кохонена.

# Список литературы

1. Методические указания «Изучение методов интеллектуального анализ данных: компонентный анализ, кластерный анализ» к лабораторному практикуму по дисциплине «Интеллектуальные технологии и представление знаний» / Уфимск. roc. авиац. техн. ун-т; Сост.: Б.Г.Ильясов, Е.А.Макарова, Э.Р.Габдуллина – Уфа, 2007.-30 с.

2. Методические указания «Изучение алгоритмов построения деревьев решений в системах *SEE/5* и *TREE* *ANALYZER*» к лабораторному практикуму по дисциплине «Интеллектуальные технологии и представление знаний» / Уфимск. roc. авиац. техн. ун-т; Сост.: Б.Г.Ильясов, Е.А.Макарова, Э.Р.Габдуллина – Уфа, 2007.-29 с.

3. Методические указания к курсовому проектированию по дисциплине «Моделирование ИАС» дли специальности АСОИУ / Уфимск. roc. авиац. техн. ун-т; Сост.: Е.А.Макарова. – Уфа, 2001.

4. Базы знаний интеллектуальных систем / Т.А.Гаврилова, В.Ф.Хорошевский – СПб: Питер, 2000 – 384 с.

5. Тельнов Ю.Ф. Интеллектуальные информационные системы в экономике. – Уч. пособие. – М.: Синтег, 1998. – 216 с.

6. Статические и динамические экспертные системы: Учеб. пособие / Э.В.Попов, И.Б.Фоминых, Е.Б.Кисель, М.Д.Шапот. – М.: Финансы и статистика, 1996. – 320 с.: ил.

7. Основные показатели здоровья матери и ребенка, деятельность службы охраны детства и родовспоможения в Российской Федерации за 2019 год: Статистический сборник / ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» !