



Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Московский государственный технический университет  
имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)

---

ФАКУЛЬТЕТ ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

КАФЕДРА КОМПЬЮТЕРНЫЕ СИСТЕМЫ И СЕТИ (ИУ6)

НАПРАВЛЕНИЕ ПОДГОТОВКИ 09.04.01 Информатика и вычислительная техника

МАГИСТЕРСКАЯ ПРОГРАММА 09.04.01/07 Интеллектуальные системы анализа,  
обработки и интерпретации больших данных

## О Т Ч Е Т

### по лабораторной работе № 2

**Название:** Выявление логических закономерностей по данным мониторинга

**Дисциплина:** Дистанционный мониторинг сложных систем и процессов

Студент

ИУ6-12М

(Группа)

\_\_\_\_\_  
(Подпись,  
дата)

Д.С. Каткова

\_\_\_\_\_  
(И.О. Фамилия)

Преподаватель

\_\_\_\_\_  
(Подпись,  
дата)

Ю.А. Вишневская

\_\_\_\_\_  
(И.О. Фамилия)

Москва, 2023

**Цель работы** – изучение способов выявления закономерностей в разнородных данных.

**Задание.** Необходимо проанализировать собранные данные, определить и закодировать информационные признаки, выбрать метод и выявить логические закономерности с его помощью:

- 1) найти закономерности группирования изображений лиц людей;
- 2) определить, чем лица разных классов отличаются друг от друга и что объединяет лица одного класса.

**Исходные данные:** имеется набор данных в виде изображений лиц людей, сгруппированных в два класса (рис. 1).

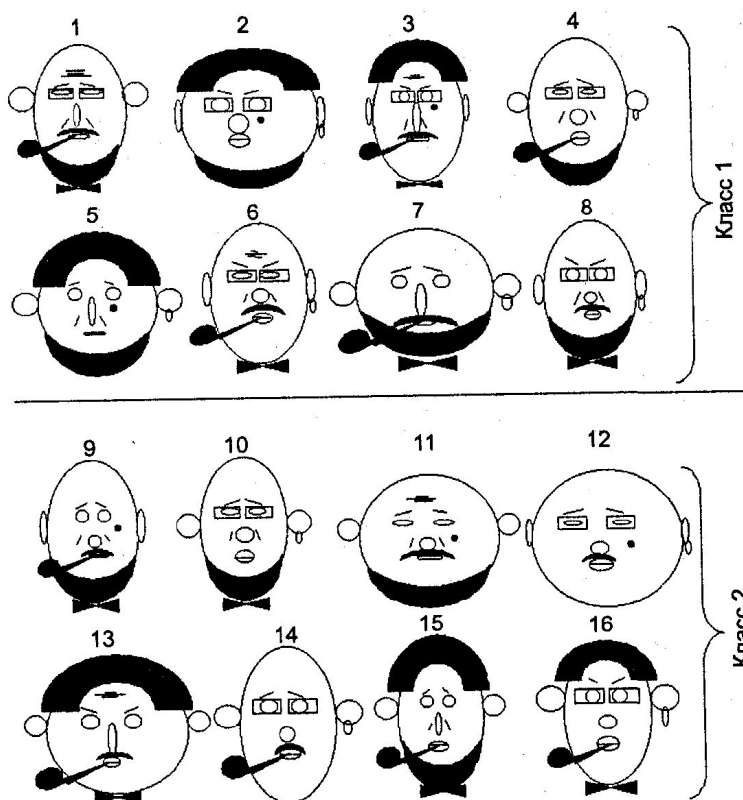


Рисунок 1 - Классы изображений лиц людей

**Методы:** ручной/визуальный метод.

**Порядок выполнения работы:**

Идентификационными признаками были выделены:

- x1 (голова): круглая – 1, овальная – 0;
- x2 (уши): оттопыренные – 1, прижатые – 0;
- x3 (нос): круглый – 1, длинный – 0;

- х4 (глаза): круглые – 1, узкие – 0;
- х5 (лоб): с морщинами – 1, без морщин – 0;
- х6 (складка): носогубная складка есть – 1, носогубной складки нет – 0;
- х7 (губы): толстые – 1, тонкие – 0;
- х8 (волосы): есть – 1, нет – 0;
- х9 (усы): есть – 1, нет – 0;
- х10 (борода): есть – 1, нет – 0;
- х11 (очки): есть – 1, нет – 0;
- х12 (родинка): родинка на щеке есть – 1, родинки на щеке нет – 0;
- х13 (бабочка): есть – 1, нет – 0;
- х14 (брови): подняты кверху – 1, опущены книзу – 0;
- х15 (серьга): есть – 1, нет – 0;
- х16 (трубка): курительная трубка есть – 1, нет – 0.

Была сформирована матрица признаков (рисунок 2).

N п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
X1	0	1	0	0	1	0	1	0	0	0	1	1	1	0	0	0
X2	1	0	0	1	1	0	1	0	0	1	1	0	1	1	1	1
X3	0	1	0	1	0	1	0	1	1	1	1	1	0	1	0	1
X4	0	1	1	0	1	0	1	1	1	0	0	0	1	1	1	1
X5	1	0	1	0	0	1	0	0	0	0	1	1	1	0	0	0
X6	1	0	1	1	1	1	0	1	1	1	1	0	0	0	1	0
X7	0	1	0	1	0	1	0	1	0	1	0	1	1	1	1	1
X8	0	1	1	0	1	0	0	0	0	0	0	0	1	0	1	1
X9	1	0	1	0	0	1	1	1	1	0	1	1	1	1	0	0
X10	1	1	0	1	1	0	1	1	1	1	1	0	0	0	1	0
X11	1	1	1	1	0	1	0	1	0	1	0	1	0	1	0	1
X12	0	1	1	0	1	0	0	0	1	0	1	1	0	0	0	0
X13	1	0	1	0	0	1	1	1	1	1	0	0	1	0	1	1
X14	1	0	0	1	1	0	1	0	1	1	1	1	0	1	1	0
X15	0	1	0	1	1	1	1	1	0	1	0	1	0	1	0	1
X16	1	0	1	1	0	1	1	0	1	0	0	0	1	1	1	1

Рисунок 2 - Матрица бинарных признаков

Для решения данной задачи будет использовано дерево решений - это метод представления решающих правил в иерархической структуре, состоящей из элементов двух типов — узлов и листьев. Узлы представляют собой условия, основанные на значениях признаков, а ветви - возможные результаты проверки условий. Каждый лист дерева определяет решение для попавших в него примеров. Для дерева классификации это класс, ассоциированный с узлом.

Определим признак с наибольшей дискриминирующей силой.  
Результаты определения представлены на рисунке 3.

	C1	C2
X1	3	3
X2	4	6
X3	4	6
X4	5	5
X5	3	3
X6	6	4
X7	4	6
X8	3	3
X9	5	5
X10	6	4
X11	6	4
X12	3	3
X13	5	5
X14	4	6
X15	6	4
X16	5	5

Рисунок 3 – Признаки с наибольшей дискриминирующей силой

Таким образом, максимальной и одинаковой силой обладают признаки X2, X3, X6, X7, X10, X11, X14, X15.

В качестве первого признака выбирается – X7. Описание результатов с данными, соответствующими ветви X7=0, приведены на рисунке 4.

N п/п	1	3	5	7	9	11	C1	C2
X1	0	0	1	1	0	1	2	1
X2	1	0	1	1	0	1	3	1
X3	0	0	0	0	1	1	0	2
X4	0	1	1	1	1	0	3	1
X5	1	1	0	0	0	1	2	1
X6	1	1	1	0	1	1	3	2
X8	0	1	1	0	0	0	2	0
X9	1	1	0	1	1	1	3	2
X10	1	0	1	1	1	1	3	2
X11	1	1	0	0	0	0	2	0
X12	0	1	1	0	1	1	2	2
X13	1	1	0	1	1	0	3	1
X14	1	0	1	1	1	1	3	2
X15	0	0	1	1	0	0	2	0
X16	1	1	0	1	1	0	3	1

Рисунок 4 – Данные, соответствующие ветви X7=0

Под ветвь  $X_7=0$  попадают лица с номерами 1, 3, 5, 7, 9, 11.

Далее в качестве признака выбирается –  $X_3=1$  (рисунок 5).

<b>N п/п</b>	<b>9</b>	<b>11</b>
<b>X1</b>	0	1
<b>X2</b>	0	1
<b>X4</b>	1	0
<b>X5</b>	0	1
<b>X6</b>	1	1
<b>X8</b>	0	0
<b>X9</b>	1	1
<b>X10</b>	1	1
<b>X11</b>	0	0
<b>X12</b>	1	1
<b>X13</b>	1	0
<b>X14</b>	1	1
<b>X15</b>	0	0
<b>X16</b>	1	0

Рисунок 5 – Данные, соответствующие ветви  $X_3=1$

Под ветвь  $X_7=0$  и  $X_3=1$  подходят лица с номерами 9 и 11 (класс 2).

Данные, соответствующие ветви  $X_{10}=0$ , показаны на рисунке 6.

<b>N п/п</b>	<b>1</b>	<b>3</b>	<b>5</b>	<b>7</b>
<b>X1</b>	0	0	1	1
<b>X2</b>	1	0	1	1
<b>X4</b>	0	1	1	1
<b>X5</b>	1	1	0	0
<b>X6</b>	1	1	1	0
<b>X8</b>	0	1	1	0
<b>X9</b>	1	1	0	1
<b>X10</b>	1	0	1	1
<b>X11</b>	1	1	0	0
<b>X12</b>	0	1	1	0
<b>X13</b>	1	1	0	1
<b>X14</b>	1	0	1	1
<b>X15</b>	0	0	1	1
<b>X16</b>	1	1	0	1

Рисунок 6 – Данные, соответствующие ветви  $X_3=0$

Под ветвь  $X_7=0$  и  $X_3=0$  подходят лица с номерами 1, 3, 5, 7 (класс 1).

Аналогично рассмотрим ветвь  $X_7=1$ . Описание результатов с данными, соответствующими ветви  $X_7=1$ , приведены на рисунке 7.

N п/п	2	4	6	8	10	12	13	14	15	16	C1	C2
X1	1	0	0	0	0	1	1	0	0	0	1	2
X2	0	1	0	0	1	0	1	1	1	1	1	5
X3	1	1	1	1	1	1	0	1	0	1	4	4
X4	1	0	0	1	0	0	1	1	1	1	2	4
X5	0	0	1	0	0	1	1	0	0	0	1	2
X6	0	1	1	1	1	0	0	0	1	0	3	2
X8	1	0	0	0	0	0	1	0	1	1	1	3
X9	0	0	1	1	0	1	1	1	0	0	2	3
X10	1	1	0	1	1	0	0	0	1	0	3	2
X11	1	1	1	1	1	1	0	1	0	1	4	4
X12	1	0	0	0	0	1	0	0	0	0	1	1
X13	0	0	1	1	1	0	1	0	1	1	2	4
X14	0	1	0	0	1	1	0	1	1	0	1	4
X15	1	1	1	1	1	1	0	1	0	1	4	4
X16	0	1	1	0	0	0	1	1	1	1	2	4

Рисунок 7 – Данные, соответствующие ветви X7=1

Под ветвь X7=1 попадают лица с номерами 2, 4, 6, 8, 10, 12, 13, 14, 15, 16.

Далее в качестве признака выбирается – X2=0 (рисунок 8).

N п/п	2	6	8	12	C1	C2
X1	1	0	0	1	1	1
X3	1	1	1	1	3	1
X4	1	0	1	0	2	0
X5	0	1	0	1	1	1
X6	0	1	1	0	2	0
X8	1	0	0	0	1	0
X9	0	1	1	1	2	1
X10	1	0	1	0	2	0
X11	1	1	1	1	3	1
X12	1	0	0	1	1	1
X13	0	1	1	0	2	0
X14	0	0	0	1	0	1
X15	1	1	1	1	3	1
X16	0	1	0	0	1	0

Рисунок 8 – Данные, соответствующие ветви X2=0

Под ветвь X7=1 и X2=0 подходят лица с номерами 2, 6, 8, 12.

Далее в качестве признака выбирается – X14=0 (рисунок 9).



<b>N п/п</b>	<b>2</b>	<b>6</b>	<b>8</b>
<b>X1</b>	1	0	0
<b>X3</b>	1	1	1
<b>X4</b>	1	0	1
<b>X5</b>	0	1	0
<b>X6</b>	0	1	1
<b>X8</b>	1	0	0
<b>X9</b>	0	1	1
<b>X10</b>	1	0	1
<b>X11</b>	1	1	1
<b>X12</b>	1	0	0
<b>X13</b>	0	1	1
<b>X15</b>	1	1	1
<b>X16</b>	0	1	0

Рисунок 10 – Данные, соответствующие ветви  $X_{14}=0$

Под ветвь  $X_7=1$ ,  $X_2=0$  и  $X_{14}=0$  подходят лица с номерами 2, 6, 8 (класс 1).

Данные, соответствующие ветви  $X_{14}=1$ , показаны на рисунке 9.

<b>N п/п</b>	<b>12</b>
<b>X1</b>	1
<b>X3</b>	1
<b>X4</b>	0
<b>X5</b>	1
<b>X6</b>	0
<b>X8</b>	0
<b>X9</b>	1
<b>X10</b>	0
<b>X11</b>	1
<b>X12</b>	1
<b>X13</b>	0
<b>X15</b>	1
<b>X16</b>	0

Рисунок 10 – Данные, соответствующие ветви  $X_2=0$

Под ветвь  $X_7=1$ ,  $X_2=0$  и  $X_{14}=1$  подходит лицо с номером 12 (класс 2).

Далее в качестве признака выбирается  $X_2=1$  (рисунок 11).

N п/п	4	10	13	14	15	16	C1	C2
X1	0	0	1	0	0	0	0	1
X3	1	1	0	1	0	1	1	3
X4	0	0	1	1	1	1	0	4
X5	0	0	1	0	0	0	0	1
X6	1	1	0	0	1	0	1	2
X8	0	0	1	0	1	1	0	3
X9	0	0	1	1	0	0	0	2
X10	1	1	0	0	1	0	1	2
X11	1	1	0	1	0	1	1	3
X12	0	0	0	0	0	0	0	0
X13	0	1	1	0	1	1	0	4
X14	1	1	0	1	1	0	1	3
X15	1	1	0	1	0	1	1	3
X16	1	0	1	1	1	1	1	4

Рисунок 11 – Данные, соответствующие ветви X2=1

Под ветвь X7=1 и X2=1 подходят лица с номерами 4, 10, 13, 14, 15, 16.

Данные, соответствующие ветви X4=1, показаны на рисунке 12.

N п/п	13	14	15	16
X1	1	0	0	0
X3	0	1	0	1
X4	1	1	1	1
X5	1	0	0	0
X6	0	0	1	0
X8	1	0	1	1
X9	1	1	0	0
X10	0	0	1	0
X11	0	1	0	1
X12	0	0	0	0
X13	1	0	1	1
X14	0	1	1	0
X15	0	1	0	1
X16	1	1	1	1

Рисунок 12 – Данные, соответствующие ветви X4=1

Под ветвь X7=1, X2=1 и X4=1 подходят лица с номерами 13, 14, 15, 16 (класс 2).

Данные, соответствующие ветви X4=0, показаны на рисунке 13.



N п/п	4	10	C1	C2
X1	0	0	0	0
X3	1	1	1	1
X4	0	0	0	0
X5	0	0	0	0
X6	1	1	1	1
X8	0	0	0	0
X9	0	0	0	0
X10	1	1	1	1
X11	1	1	1	1
X12	0	0	0	0
X13	0	1	0	1
X14	1	1	1	1
X15	1	1	1	1
X16	1	0	1	0

Рисунок 13 – Данные, соответствующие ветви X4=0

Далее под ветвь X7=1, X2=1, X4=1 и X13=0 подходит лицо с номером 4 (класс 1). А под ветвь X7=1, X2=1, X4=1 и X13=1 подходит лицо с номером 10 (класс 2).

Таким образом, было получено дерево решений. Дерево решений представлено на рисунке 14.

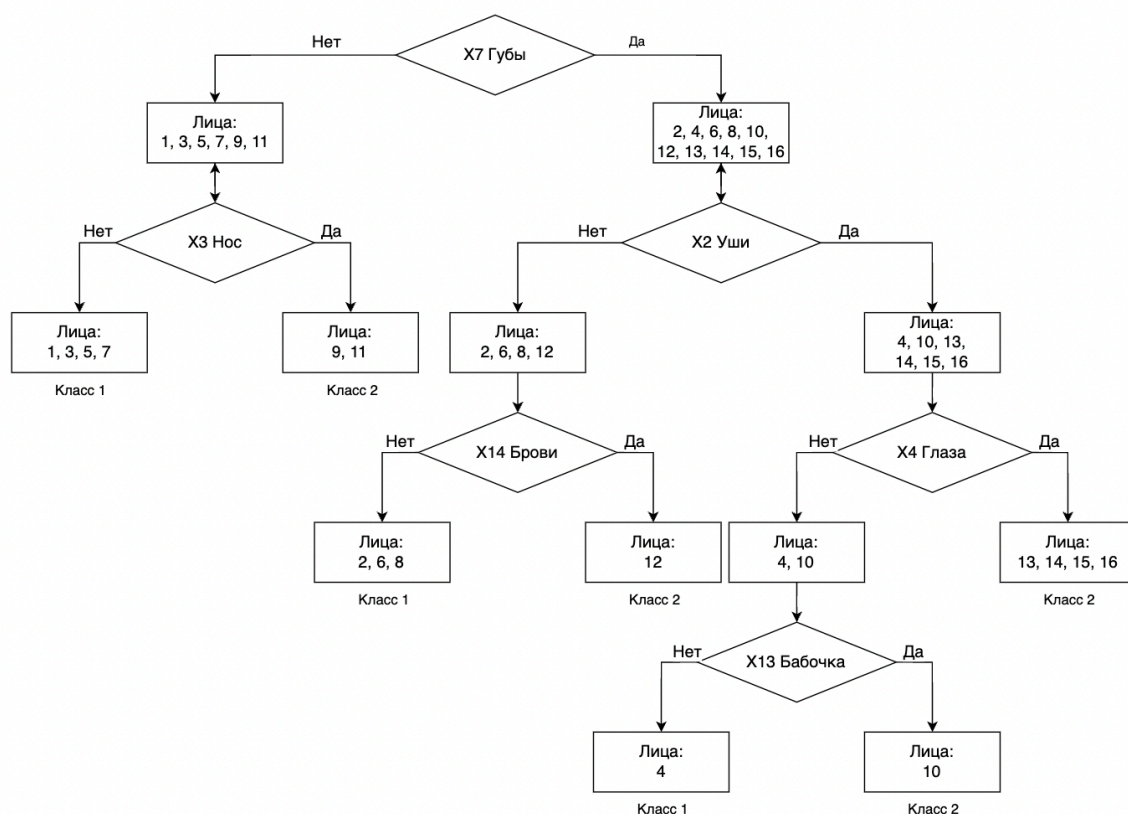


Рисунок 14 – Дерево решений

Из дерева решений видно, что для разделения лиц на классы необходимо 7 параметров:

- 1) губы;
- 2) нос;
- 3) уши;
- 4) брови;
- 5) глаза;
- 6) бабочка.

По данной схеме можно сформулировать следующие правила:

- 1) ЕСЛИ {(губы тонкие) и (нос круглый)} ТО (Класс 1).
- 2) ЕСЛИ {(губы тонкие) и (нос длинный)} ТО (Класс 2).
- 3) ЕСЛИ {(губы толстые) и (уши прижатые) и (брови опущены к низу)} ТО (Класс 1).
- 4) ЕСЛИ {(губы толстые) и (уши прижатые) и (брови подняты кверху)} ТО (Класс 2).
- 5) ЕСЛИ {(губы толстые) и (уши оттопыренные) и (глаза узкие) и (бабочки нет)} ТО (Класс 1).
- 6) ЕСЛИ {(губы толстые) и (уши оттопыренные) и (глаза узкие) и (бабочка есть)} ТО (Класс 2).
- 7) ЕСЛИ {(губы толстые) и (уши оттопыренные) и (глаза круглые)} ТО (Класс 2).

**Вывод:** в результате данной лабораторной работы был проведен системный анализ данных, изучены различные способы выявления закономерностей в данных и реализован один из них.

## **Контрольные вопросы**

**1. Что понимается под закономерностями в данных? Приведите примеры типовых закономерностей.**

Под закономерностью в данных понимается взаимосвязь данных между собой. Пример: определить, какой товар предложить пользователю на основе просмотренных товаров.

**2. Поясните основные подходы выявления и анализа закономерностей внутри данных.**

– Кластерный анализ: метод, позволяющий разбивать данные на группы (кластеры) схожих объектов на основе их характеристик.

– Факторный анализ: метод, который помогает выделить основные факторы, влияющие на данные, и установить зависимости между ними.

– Ассоциативный анализ: метод, позволяющий находить связи между различными атрибутами данных и исследовать их влияние друг на друга.

– Машинное обучение: метод, который использует алгоритмы для изучения больших объемов данных и нахождения скрытых закономерностей.

– Регрессионный анализ: метод, используемый для выявления зависимости между независимыми и зависимыми переменными в данных.

– Методы искусственного интеллекта, такие как нейронные сети или генетические алгоритмы.

– Алгоритмы обнаружения аномалий, которые используются для выявления необычных или неожиданных паттернов в данных.

– Техники визуализации данных, которые позволяют исследовать данные в графическом формате и выявлять скрытые закономерности.

**3. Укажите и поясните основные статистические методы для выявления закономерностей в данных.**

– регрессионный, дисперсионный и корреляционный анализ;

– методы анализа в конкретной предметной области, базирующиеся на эмпирических моделях;

- нейросетевые алгоритмы, идея которых основана на аналогии с функционированием нервной ткани и заключается в том, что исходные параметры рассматриваются как сигналы, преобразующиеся в соответствии с имеющимися связями между «нейронами», а в качестве ответа, являющегося результатом анализа, рассматривается отклик всей сети на исходные данные;
- деревья решений;
- кластерные модели;
- алгоритмы ограниченного перебора, вычисляющие частоты комбинаций простых логических событий в подгруппах данных.

**4. Укажите интеллектуальные методы, применяемые для анализа больших данных.**

- машинное обучение;
- дата майнинг;
- нейросети;
- имитационные модели;
- предикативный и статистический анализ.

**5. С какой целью проводится кодирование информационных признаков?**

Чтобы можно было представить информацию в формате матрицы 0 и 1, и использовать её для математической обработки с помощью машинного обучения посредством написания кода. Или для простого более наглядного, упорядоченного представления информации.

**6. Как можно определить логические закономерности в данных?**

Полученную информацию можно представить графически, использовать графы или закодировать и отсортировать в таблице. После чего либо ручным анализом (например деревья), либо с помощью компьютерных программ, выявляющим закономерности в закодированной информации.

**7. С какой целью проводится предварительная обработка данных при мониторинге? Что она включает?**

Предварительная обработка данных проводится для повышения качества данных. Она включает:

- очистку данных, которая направлена на повышение качества данных за счет присваивания пропущенных значений и удаления выбросов;
- сокращение объема данных, которое уменьшает объем данных и, следовательно, снижает связанные с ними вычислительные мощности;
- масштабирование данных – направлено на преобразование исходных данных в аналогичные диапазоны для прогнозирования;
- преобразование, целью которого является организация исходных данных в подходящие форматы для различных алгоритмов анализа данных;
- разделение, которое делит весь набор данных на различные подмножества для более глубокого анализа.