

Министерство науки и высшего образования Российской Федерации
Муромский институт
федерального государственного бюджетного образовательного учреждения
высшего образования
**«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»
(МИВлГУ)**

Факультет _____ ИТ _____

Кафедра _____ ИС _____

ЛАБОРАТОРНАЯ РАБОТА №2

по _____ Интеллектуальные системы и технологии _____

Тема _____ Подготовка данных для машинного обучения и оценка
качества обученного классификатора

Руководитель

Варламов А.Д.
(фамилия, инициалы)

(подпись) (дата)

Студент ИС-117
(группа)

Минеев Р.Р.
(фамилия, инициалы)

(подпись) (дата)

Муром 2021

Лабораторная работа №2.

Тема: Подготовка данных для машинного обучения и оценка качества обученного классификатора.

Цель работы: Изучить методы анализа и трансформации данных, применяемых в машинном обучении.

Задание на работу:

- Создать датасет
- Оценить качество
- Обучить классификатор
- Оценить качество обученного классификатора

Результат работы:

Создание датасета

```
with open('../dataset_StudentsOtchislenie.csv', 'w') as dataset:
    dataset.write('ii;isis;pis;iis;otchislenie\n')
    for _ in range(10000):
        ii = random.randint(0, ii_max)
        isis = random.randrange(0, isis_max)
        pis = random.randrange(0, pis_max)
        iis = random.randrange(0, iis_max)
        total_score = sum([ii, isis, pis, iis])
        otchislenie = True if total_score <= max_score / 2 else False
        string = f'{ii};{isis};{pis};{iis};{int(otchislenie)}\n'
        dataset.write(string)
```

ii	isis	pis	iis	otchislenie
8	10	16	13	0
10	5	5	12	1
0	8	14	0	1
15	23	13	18	0
6	18	7	16	0
13	18	17	18	0
7	7	12	18	0
7	13	16	11	0
16	6	6	1	1
8	2	16	10	1
19	23	17	2	0
6	21	18	18	0
5	0	12	6	1

					МИВУ 09.03.02-02.005			
Изм	Лист	№ докум.	Подп.	Дата	Лабораторная работа №2 Подготовка данных для машинного обучения и оценка качества обученного классификатора.	Литера	Лист	Листов
Студент	Минеев Р. Р.			18.03.		У	2	3
Руков.	Варламов А.Д.					МИ ВлГУ ИС-117		
Конс								
Н.контр.								
Утв.								

Загрузка датасета и выявление дубликатов и противоречий:

Дубликаты		Противоречия		Входные поля				Выходные поля
Признак	Группа	Признак	Группа	ii	isis	pis	iis	otchislenie
<input type="checkbox"/>		<input type="checkbox"/>		13	20	14	17	<input type="checkbox"/>
<input type="checkbox"/>		<input type="checkbox"/>		19	6	12	15	<input type="checkbox"/>
<input type="checkbox"/>		<input type="checkbox"/>		14	2	2	12	<input checked="" type="checkbox"/>
<input type="checkbox"/>		<input type="checkbox"/>		20	17	9	6	<input type="checkbox"/>
<input checked="" type="checkbox"/>	23	<input type="checkbox"/>		2	9	3	4	<input checked="" type="checkbox"/>
<input type="checkbox"/>		<input type="checkbox"/>		7	21	8	17	<input type="checkbox"/>
<input type="checkbox"/>		<input type="checkbox"/>		0	7	2	5	<input checked="" type="checkbox"/>
<input type="checkbox"/>		<input type="checkbox"/>		6	5	18	12	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	174	<input type="checkbox"/>		14	22	12	0	<input type="checkbox"/>
<input type="checkbox"/>		<input type="checkbox"/>		11	18	11	6	<input type="checkbox"/>
<input checked="" type="checkbox"/>	188	<input type="checkbox"/>		16	4	9	18	<input type="checkbox"/>
<input type="checkbox"/>		<input type="checkbox"/>		13	6	4	18	<input checked="" type="checkbox"/>
<input type="checkbox"/>		<input type="checkbox"/>		15	14	16	18	<input type="checkbox"/>
<input checked="" type="checkbox"/>	10	<input type="checkbox"/>		1	2	12	15	<input checked="" type="checkbox"/>
<input type="checkbox"/>		<input type="checkbox"/>		20	7	13	10	<input type="checkbox"/>

Проводим корреляционный анализ

Корреляционный анализ

Выбор значащих и исключение незначащих факторов

Входные поля	Корреляция с выходными полями	
	otchislenie	
<input checked="" type="checkbox"/> ii		-0,398
<input checked="" type="checkbox"/> isis		-0,447
<input checked="" type="checkbox"/> pis		-0,376
<input checked="" type="checkbox"/> iis		-0,367

Делаем вывод о том, что все признаки в равной доле важны для обучения

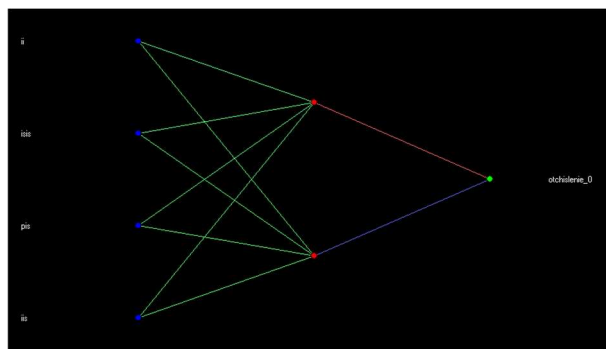
Проводим факторный анализ

Факторный анализ

Уточнение числа выделяемых факторов в соответствии с желаемой долей воспроизводимой ими дисперсии (правый столбец)

Главные компоненты	Собственное значение	Вклад в результат	Суммарный вклад
<input checked="" type="checkbox"/> Значение 1	1,787	35,7302 %	35,7302 %
<input checked="" type="checkbox"/> Значение 2	1,021	20,4212 %	56,1514 %
<input checked="" type="checkbox"/> Значение 3	1,012	20,2322 %	76,3835 %
<input checked="" type="checkbox"/> Значение 4	0,988	19,7603 %	96,1439 %
<input type="checkbox"/> Значение 5	0,193	03,8561 %	

Обучаем нейросеть



Поле	Значение
Входные	
12 ii	20
12 isis	20
12 pis	10
12 iis	2
Выходные	
0/1 otchislenie	False

Вывод: В данной лабораторной работе были получены навыки в построении и обучении нейронных сетей при помощи программ-нейроимитаторов (нейросимуляторов), а также проведён анализ данных для сокращения обучающей выборки.

Изм.	Лист	№ докум.	Подп.	Дата