МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ

УНИВЕРСИТЕТ им. Р.Е.АЛЕКСЕЕВА

Институт радиоэлектроники и информационных технологий

Кафедра информатики и систем управления

ОТЧЕТ

по лабораторной работе № 1

«Предварительная обработка данных»

по дисциплине

Машинное обучение

РУКОВОДИТЕЛЬ:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Гай В.Е.

(подпись) (фамилия, и.,о.)

СТУДЕНТ:

Халеев А.А.

(подпись) (фамилия, и.,о.)

21-ВМз

(шифр группы)

Работа защищена «\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_

С оценкой \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Нижний Новгород 2024

**Задание**

Titanic

1. Определите количество мужчин и женщин, которые ехали на корабле.  
2. Определите какой части пассажиров удалось выжить. Посчитайте долю выживших пассажиров.  
3. Какую долю пассажиры первого класса составляли среди всех пассажиров?  
4. Какого возраста были пассажиры? Посчитайте среднее и медиану возраста пассажиров.  
5. Коррелируют ли число братьев/сестер с числом родителей/детей? Посчитайте корреляцию  
Пирсона между признаками SibSp и Parch. Оцените значение p-value. Постройте плотность  
распределения признаков SibSp и Parch.  
6. Какое самое популярное женское имя на корабле? Извлеките из полного имени пассажира  
(колонка Name) его личное имя (First Name). Попробуйте вручную разобрать несколько значений  
столбца Name и выработать правило для извлечения имен, а также разделения их на женские и  
мужские.  
7. Коррелирует ли класс, которым ехал пассажир, с выживаемостью?  
8. Визуализируйте гистограммы возраста для выживших и не выживших пассажиров. Сделайте  
выводы. Отобразите данные на одном и нескольких графиках  
9. Визуализируйте гистограммы возраста для выживших и не выживших пассажиров по классам.  
Сделайте выводы.  
10. Постройте столбчатую диаграмму количества людей: мужчины, женщины, дети.

Описание значений используемых полей (признаков) в исследуемом наборе данных:

1. pclass - класс билета, принимает 3 возможных значения

1 = 1й, 2 = 2й, 3 = 3й

1. survived – идентификатор выживания пассажира, принимает 2 возможных значения

0 = не выжил, 1 = выжил

1. name
2. sex
3. age
4. sibsp
5. parch
6. ticket
7. fare
8. cabin
9. embarked
10. boat
11. body
12. home.dest