Содержание

[«heart» 2](#_Toc107926896)

[«cortex nuclear» 3](#_Toc107926897)

[«Data\_User\_Modeling\_Dataset\_Hamdi Tolga KAHRAMAN» 4](#_Toc107926898)

[“Stars” 6](#_Toc107926899)

[Датасет «Absenteeism at work Data Set» 7](#_Toc107926900)

## «heart»

***Источник:*** <https://www.kaggle.com/datasets/pritsheta/heart-attack>

***Описание:***

Возраст(vozrast) – возраст обследуемого;

Пол(pol) – пол обследуемого;

Боль в груди(grud) – жалоба обследуемого на боль в области груди;

Давление покой(davlen) – кровяное давление обследуемого в состоянии покоя;

Холестерин(holesterin) – показание уровня холестерина обследуемого;

Сахар выше нормы(sahar) – есть ли у обследуемого уровень сахара выше нормы;

Экг покой(ekg pokoi) – уровень экг в состоянии покоя у обследуемого;

Пульс макс(pulse max) – максимальный пульс обследуемого в состоянии покоя;

Ангина(angina) – была ли ангина у обследуемого;

Талия(taliya) – размер талии у обследуемого;  
Output – был ли сердечный приступ у обследуемого(target).

## «cortex nuclear»

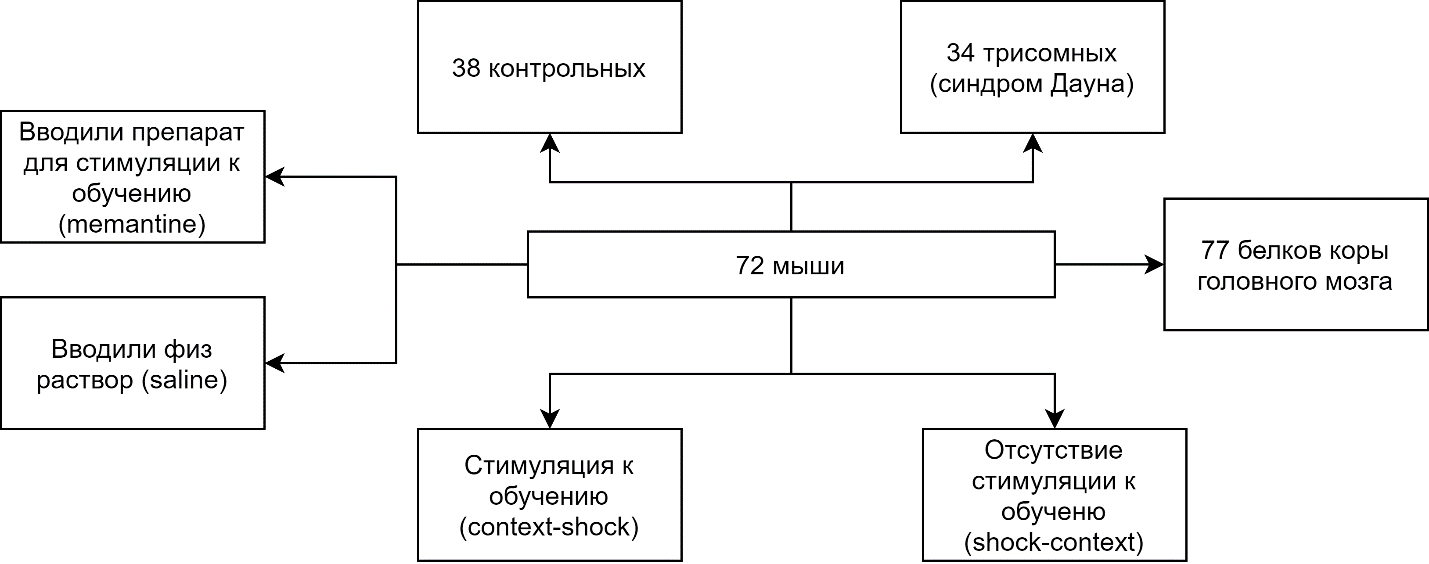
***Источник***: <https://archive.ics.uci.edu/ml/datasets/Mice+Protein+Expression>

***Описание:***

Набор данных состоит из 1080 строк и 82 столбца. Каждая строка — это индивидуальная мышь

* (1-ый столбец id) – индентификатор мыши;
* (2-78 столбцы) - значения белков;
* (79 столбец – «genotype» TARGET) - здоровая мышь или нет;
* (80 столбец – «treatment») - препарат лечения;
* (81 столбец – «behavior») - тип поведения;
* (82 столбец «class») - это сокращённая запись: genotype, behavior и treatment;

На рисунке представлена схема набора данных.



## «Data\_User\_Modeling\_Dataset\_Hamdi Tolga KAHRAMAN»

Анализируемый набор данных представляет собой информацию о результатах сдачи экзаменов по профильным и смежным дисциплинам студентов факультета электротехники и электроники университета Гази (Анкара, Турция).

Ссылка: <https://archive.ics.uci.edu/ml/datasets/User+Knowledge+Modeling>

Цель работы: провести классификацию и кластеризацию студентов на основе имеющихся данных, установить ключевые факторы влияния на успешность освоения курса.

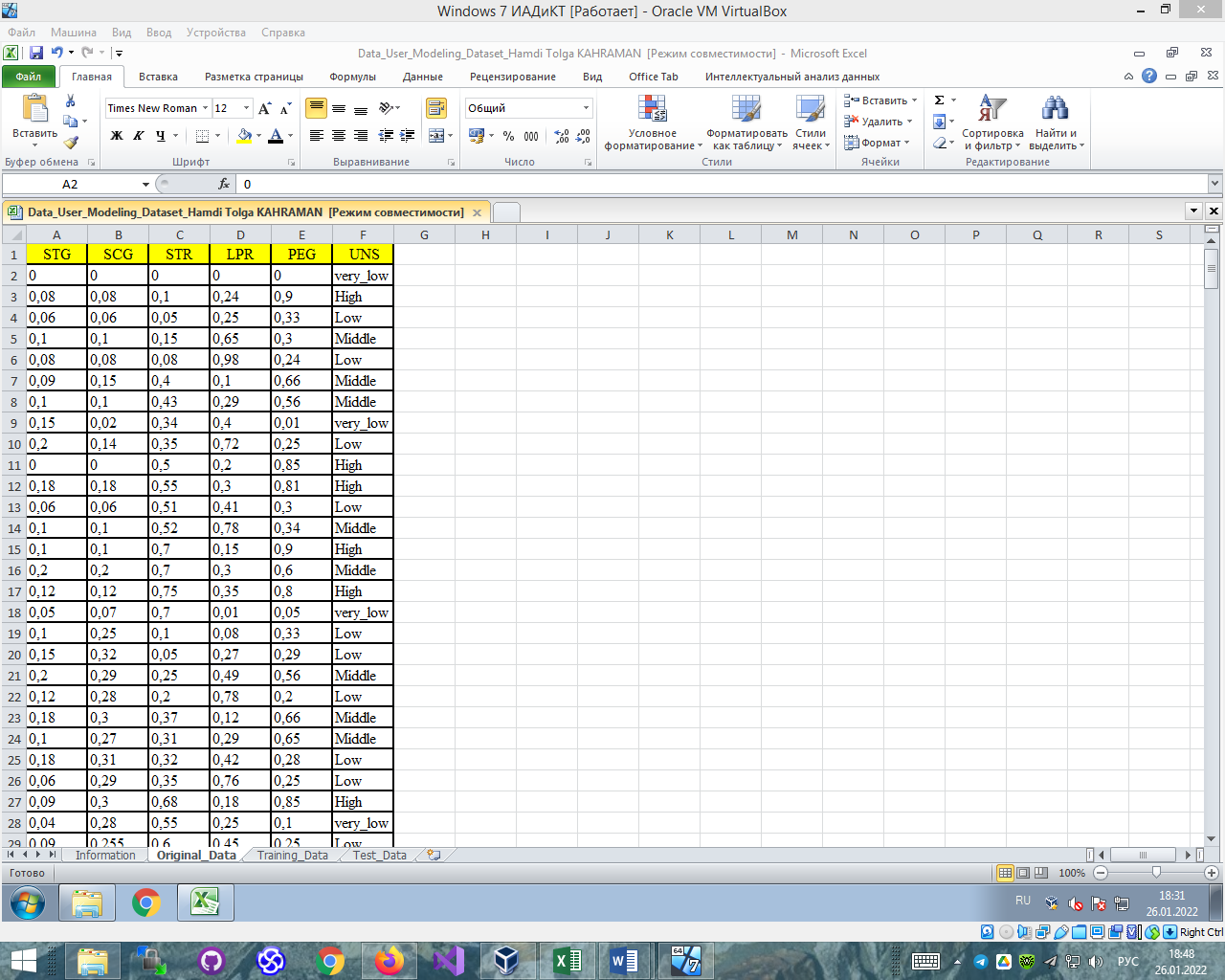


Рисунок 1 — представление данных

Пояснение обозначений:

* STG — количество эквивалентных академическим часам условных единиц, затраченных на освоение курса по профильному предмету
* SCG — количество эквивалентных академическим часам условных единиц, затраченных на подготовку к экзамену
* STR — количество эквивалентных академическим часам условных единиц, затраченных на освоение курсов по смежным дисциплинам
* LPR — результат сдачи экзамена по смежным дисциплинам (среднее арифметическое оценок)
* PEG — результат сдачи экзамена по профильному предмету
* UNS — итоговая оценка уровня знаний **(независимая переменная)**

## “Stars”

Источник: https://www.kaggle.com/datasets/deepu1109/star-dataset

Temperature – температура в Кельвинах

L - свечение

R – относительный радиус

A\_M – относительной объем

Color – цвет звезды

Spectral\_class – спектральный класс

Type – тип звезды

на основе характеристик классифицируются звезды (Type), основными характеристиками являются A\_M, L и R. Превостепенная классификация по массе и объему

# Датасет «Absenteeism at work Data Set»

Источник: https://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work

ID – индивидуальный идентификатор

ICD(от 1 до 28) – причина отсутствия на работе

Month of absence – месяц отсутствия

Day of the week(2 – понедельник, 3 – вторник, 4 – среда, 5 – четверг, 6 - пятница) – день недели

Season(1 – лето, 2 – осень, 3 – зима, 4 - весна) – время года

Fare – транспортные расходы

Distance from home to work(km) – дистанция от дома до работы в км.

Work experience(years) – стаж работы в годах

Age – возраст

Average load per day – средняя загруженность в день

Goal achieved – на сколько % выполнена цель

Disciplinary failure(1 – да, 0 - нет) – есть ли дисциплинарная неуспеваемость

Education(1 – среднее, 2 – высшее, 3 – аспирантура, 4 – магистр или доктор) – образование

Son – количество детей

Social drinker(1 – да, 0 - нет) – выпивает ли человек алкогольные напитки

Social smoker(1 – да, 0 – нет) – курит ли человек

Pet – количество домашних животных

Weight – вес

Height – рост

Body mass index – индекс массы тела

Time away from work in hours – время отсутствия на работе в часах