**МИНОБРНАУКИ РОССИИ САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ «ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)**

**Кафедра информационных систем**

**ОТЧЕТ**

**по практической работе №1**

**по дисциплине «Машинное обучение»**

**Тема: «Исследование набора данных»**

Студент гр. 2373 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Красножён Е. Д.

Преподаватель \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Татчина Я. А.

Санкт-Петербург

2024

**Цель работы:** провести статистический анализ данных выбранного датасета.

**Краткое описание датасета**

Набор данных содержит подробную информацию о 167 чемпионах из игры League of Legends. Датасет содержит строковые и числовые значения, отображающие все присущие чемпионам параметры.

В датасете представлены следующие атрибуты:

* Name — имя чемпиона League of Legends (строковый тип)
* Tags — классы чемпиона. (строковый тип)
* Role — основная роль или позиция, на которой играет чемпион. (строковый тип)
* Range type — является ли чемпион ближним или дальним боем. (строковый тип)
* Resource type — Ресурс, который чемпион может генерировать и потреблять при использовании способностей или базовой атаки. (строковый тип)
* Base HP — базовые очки здоровья чемпиона (HP) на уровне 1. (числовой тип)
* HP per lvl — количество HP, которое чемпион получает за уровень. (числовой тип)
* Base mana — базовые очки маны (MP) чемпиона на уровне 1. (числовой тип)
* Mana per lvl — количество очков маны (MP), которое чемпион получает за уровень. (числовой тип)
* Movement speed — базовая скорость передвижения чемпиона. (числовой тип)
* Base armor — базовая броня чемпиона на уровне 1. (числовой тип)
* Armor per lvl — количество брони, которую чемпион получает за уровень. (числовой тип)
* Base magic resistance — базовое сопротивление магии чемпиона на уровне 1. (числовой тип)
* Magic resistance per lvl — количество сопротивления магии, которое чемпион получает за уровень. (числовой тип)
* Attack range — максимальный радиус, с которого чемпион может атаковать. (числовой тип)
* HP regeneration — определяет количество восстановлиаемого здоровья чемпиона за 5 секунд. (числовой тип)
* HP regeneration per lvl — количество восстановливаемого здоровья за 5 секунд, которое чемпион получает за уровень. (числовой тип)
* Mana regeneration — определяет количество восстановливаемой маны чемпиона за 5 секунд. (числовой тип)
* Mana regeneration per lvl — количество восстановливаемой маны за 5 секунд, которое чемпион получает за уровень. (числовой тип)
* Attack damage — базовый урон от атаки чемпиона на уровне 1. (числовой тип)
* Attack damage per lvl — количество урона от атаки, которое чемпион получает за уровень. (числовой тип)
* Attack speed per lvl — количество скорости атаки, которое чемпион получает за уровень, (%). (числовой тип)
* Attack speed — базовая скорость атаки чемпиона (AS). (числовой тип)
* AS ratio — регулирует эффективность мультипликаторов скорости атаки из всех источников. (числовой тип)

Этот датасет был считан с помощью библиотеки pandas, и с ним была произведена дальнейшая работа.

**Определение параметров**

Наличие пропущенных значений проверялось при помощи функции библиотеки pandas pd.isna().sum(), которая выводит количество пропущенных значений для каждого атрибута. Пропущенные значения обрабатывались функцией pd.fillna(), заменяя пустые ячейки на значение ‘None’

Для каждого числового атрибута были определены среднее значение и СКО. Для этого использовались функции библиотеки numpy: np.mean() и np.std() (данные представлены в файле с программой).

Поиск выбросов был осуществлен с помощью следующего алгоритма:

1. Отсортировать данные и найти Q1 и Q3 (1-ый и 3-ый квартиль)
2. Найти межквартильный размах InterQ
3. Проверить, какие наблюдения вышли за границы

[Q1 – 1.5 \* InterQ; Q3 + 1.5 \* InterQ]

Также для каждого числового атрибута была построена гистограмма распределения значений при помощи функции библиотеки matplotlib plt.hist().

**Корреляция между параметрами**

Корреляция между атрибутами была выделена с помощью функции sns.heatmap(), выводящей корреляционную матрицу. По корреляционной матрице были определены следующие пары атрибутов с высокой корреляцией:

Положительная:

* Magic resistance per lvl и Base armor
* Magic resistance per lvl и Attack damage
* Mana regeneration per lvl и Base mana
* Mana regeneration per lvl и Mana per lvl

Отрицательная:

* Attack range и Base armor
* Attack range и Magic resistance per lvl
* Attack range и Attack damage

По выявленным парам атрибутов были созданы графики рассеивания и дополнительно график рассеивания для AS ratio и Armor per lvl как для атрибутов без корреляции.

**Вывод**

В работе был проведён статистический анализ датасета, в том числе определение средних значений, СКО и выявление выбросов.

Была определена корреляцию атрибутов, выявлены атрибуты с высокой положительной и отрицательной корреляцией и построены графики рассеивания. В связи с малым количеством данных в датасете графики рассеивания недостаточно хорошо отображают корреляцию атрибутов.

Весь код программы, гистограммы, графики рассеивания и корреляционная матрица представлены в .ipynb файле.