

Работа для 2 практики

Холкин Николай

16 сентября 2025 г.

1 Предметная область

Вселенная «Песни льда и пламени» (Игры престолов)

1.1 Гипотеза

Персонажи, появившиеся в последних главах, имеют больше шансов на выживание, чем персонажи из ранних глав.

1.2 Необходимый объем данных

$\approx 1000-1500$ сэмплов

1.3 Источники сбора данных

- Kaggle

Можно выбрать только один источник данных - выгрузку парсинга с вики. В основе вики лежит информация из книг Джорджа Мартина, каким-либо другим способом получить информацию о персонажах не получится.

1.4 Объем собранных данных

Датасет с Kaggle:

- 21 параметр
- 1946 объектов

Природа данных естественная, датасет был собран путем парсинга вики фандома, википедия заполнялась по информации из первоисточников.

1.5 План для анализа данных

1. Проверить гипотезу
2. Выявить признаки, указывающую на смерть персонажа:
 - (a) Раннее появление в серии повышает вероятность смерти?
 - (b) «Высокая» связанность с другими мёртвыми персонажами повышает вероятность смерти?

3. Выявить признаки, указывающие на выживание персонажа
 - (а) Высокое соц. положение повышает шансы выживания?
4. Обучить алгоритм, который будет предсказывать судьбу персонажа по его характеристикам