

Прикладной статистический анализ

Разработка модели прогнозирования инцидентов в авиации

Автор: Савосин А.А. группа БИВТ-21-5
Научный руководитель: Маркарян А.О.



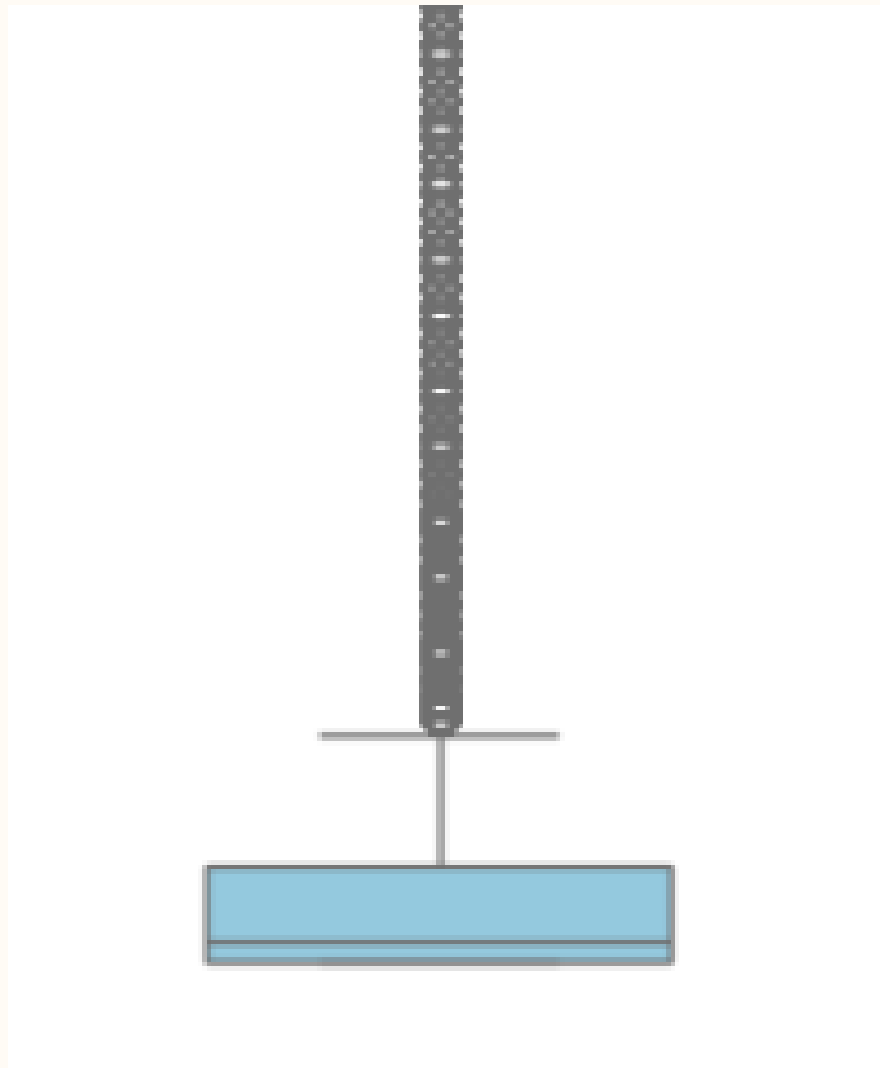
01 – Моделируемый объект

Предсказание количества летальных исходов при авиакатастрофах поможет спасательным службам заранее понимать какое кол-во инструментов и ресурсов будет необходимо для оказания скорой медицинской помощи на месте происшествия.

В данной работе
рассматривается
датасет
авиационных
инцидентов
(1918–2022 гг.)



02 – Статистические показатели



Boxplot

Среднее арифметическое – 5.5674

Дисперсия – 278.3312

Среднее квадратич. отклонение – 16.71

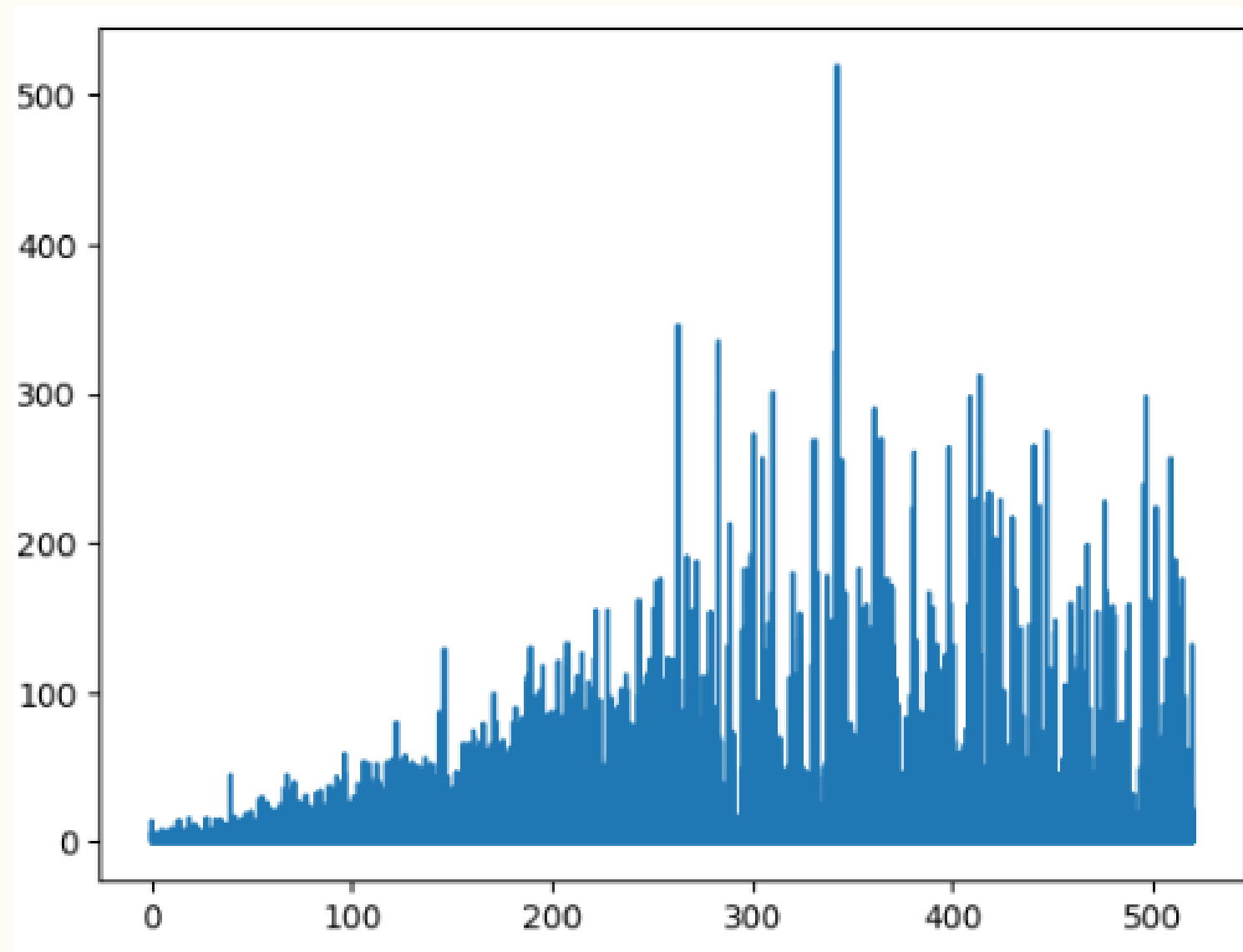
Коэффициент вариации – 300%

Размах (min 0, max 520) – 520

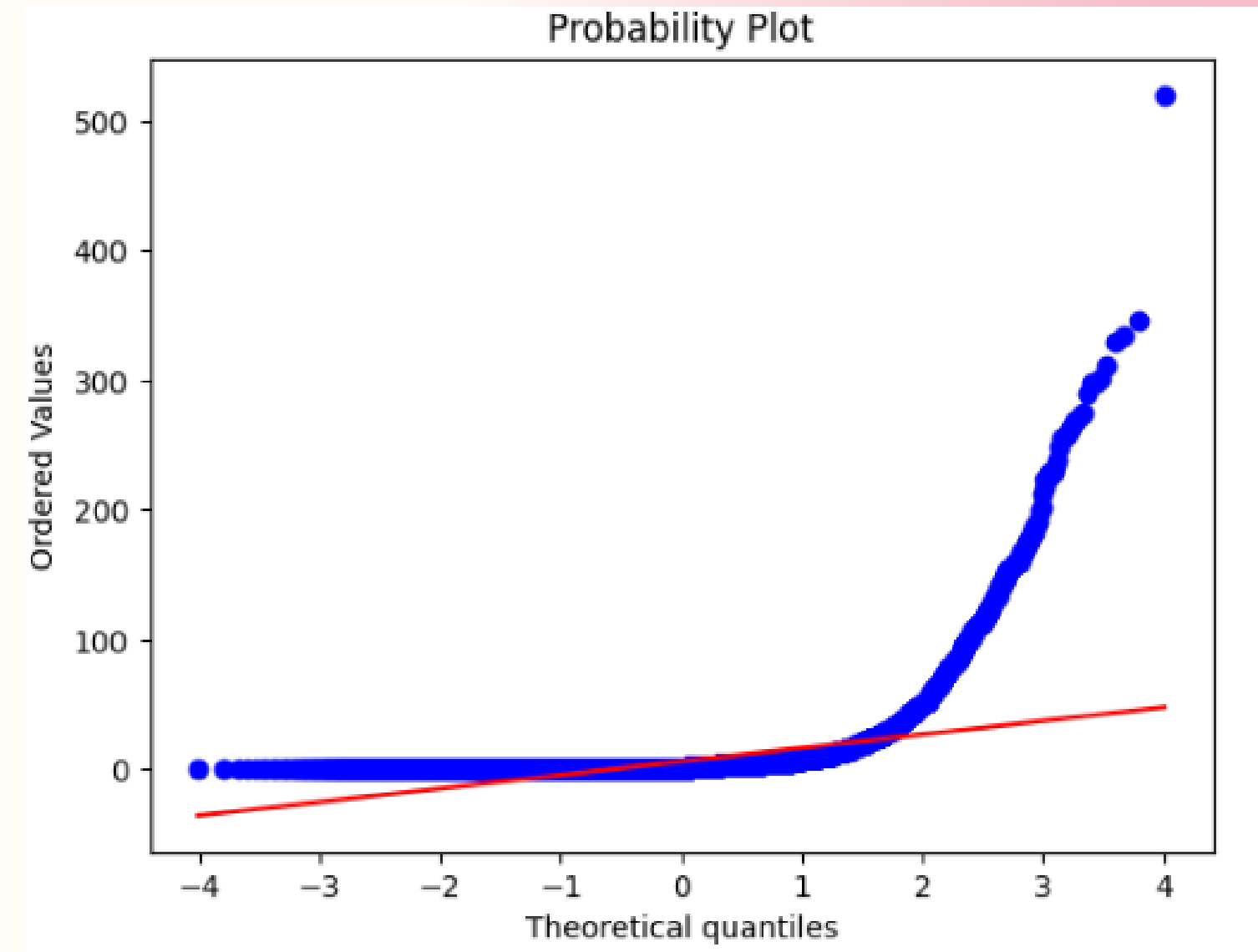
Мода с частотой 11851 – 0

Медиана – 1

02 – Статистические показатели

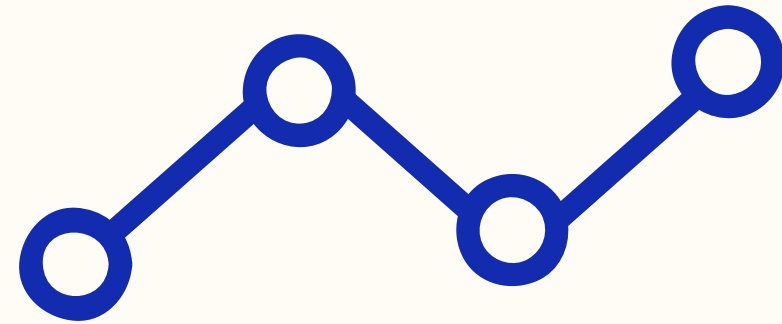


Гистограмма



QQ-plot

03 – Постановка задачи моделирования



- Предобработка признаков
- Построение регрессионной модели
- Анализ структуры модели
- Оценка эффективности модели
- Улучшение модели

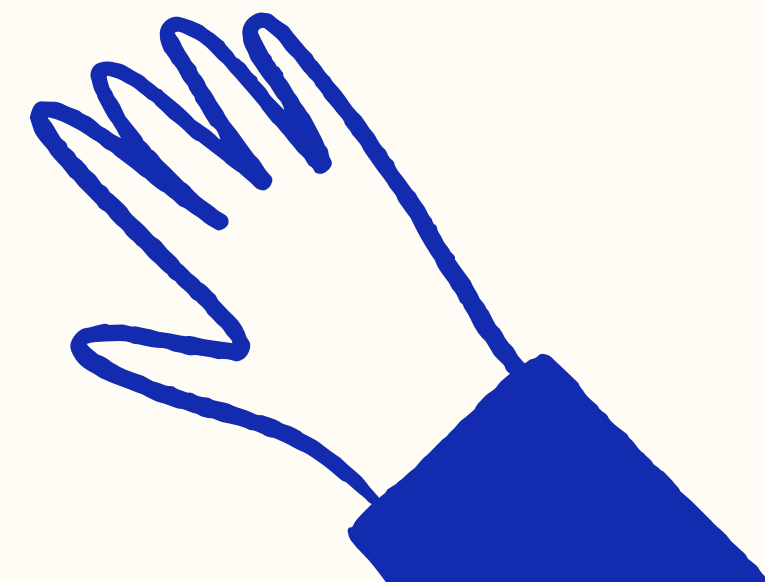
Задача заключается в предсказании потенциального количества смертей на основе данных о крушениях.



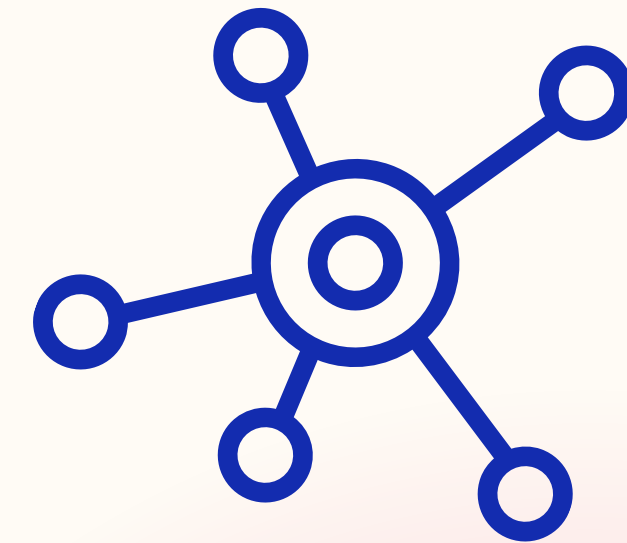
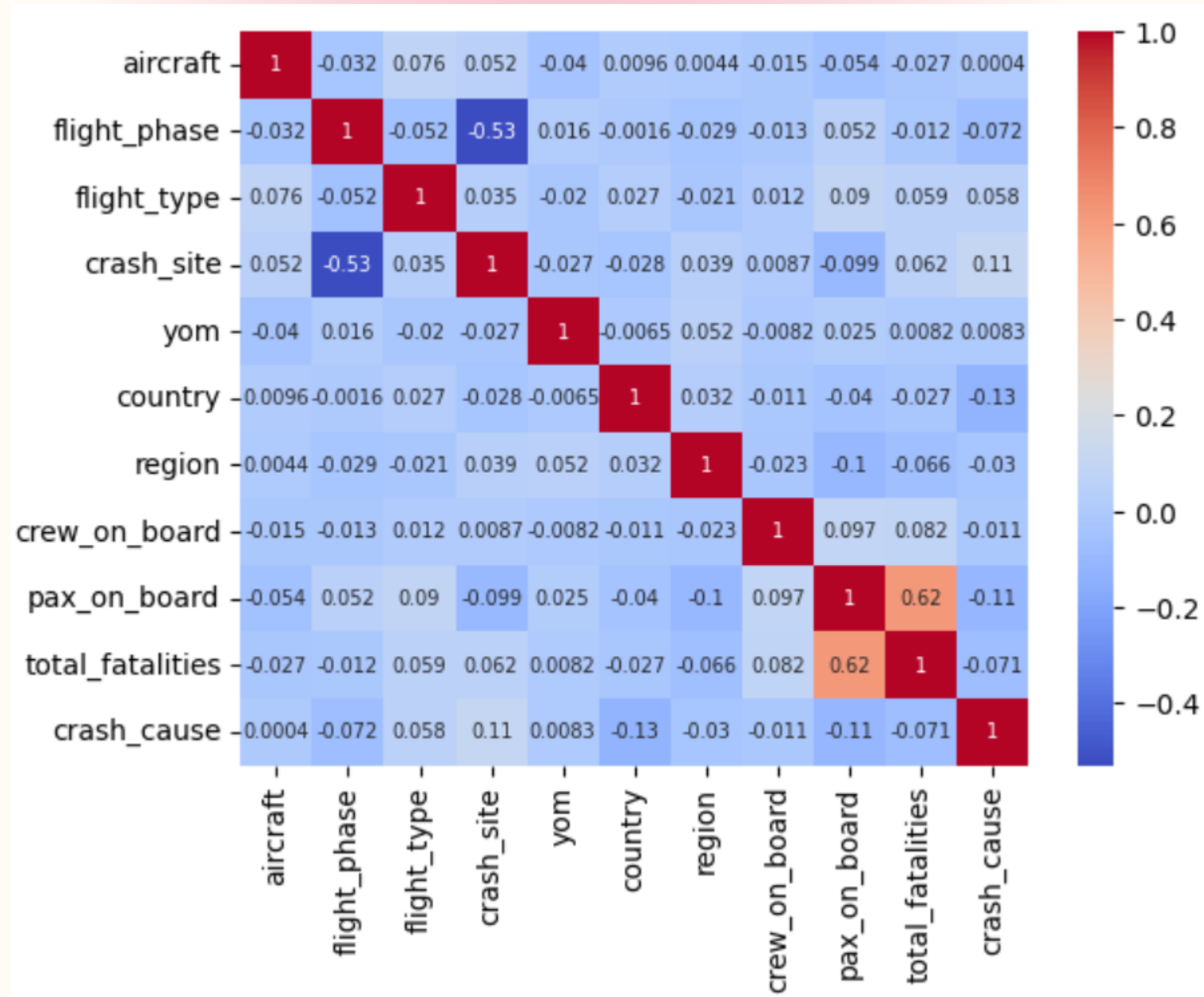
1. Aircraft – качественная, модель самолёта
2. Flight phase – качественная, стадия полёта
3. Flight type – качественная, тип полёта
4. Crash site – качественная, место крушения
5. YOM – качественная, год производства самолёта
6. Country – качественная, страна
7. Region – качественная, регион мира
8. Crew on board – количественная, кол-во персонала
9. Pax on board – количественная, кол-во пассажиров
10. Crash cause – качественная, причина крушения

Выходная количественная переменная – кол-во летальных исходов

04 – Формализация и
классификация
переменных



05 – Корреляционный анализ



Определитель матрицы
парных корреляций 0.3865

06 – Уравнение модели

Уравнение множественной
линейной регрессии

$$\begin{aligned} Y = & 0.000324 + 0.453436 * X2 - 0.001627 * X3 \\ & + 1.203601 * X4 - 0.000233 * X5 - 0.000014 * X6 \\ & - 0.049325 * X7 + 0.029225 * X8 + 0.43522 * X9 \\ & - 0.160953 * X10 \end{aligned}$$

07 – Исследование и улучшение модели

Шаговый регрессионный анализ

[flight_phase, crash_site, region, crew_on_board, pax_on_board, crash_cause]

$$\begin{aligned} Y = & -0.4498 + 0.4537 * X2 + 1.2046 * X4 \\ & - 0.0512 * X7 + 0.0293 * X8 + 0.4351 * X9 \\ & - 0.1620 * X10 \end{aligned}$$

До $R^2_{adj} = 0.39937.$

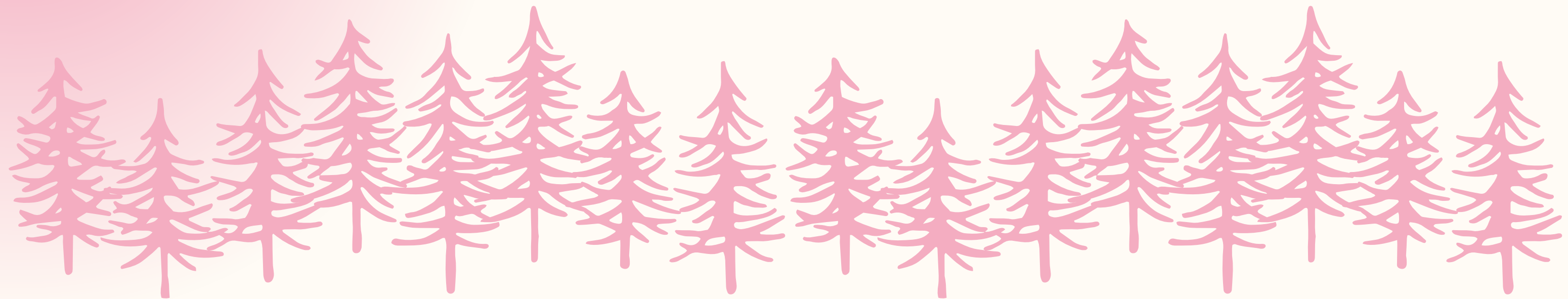
После $R^2_{adj} = 0.3994578$

07 – Исследование и улучшение модели

Машинное обучение. Метод случайного леса

```
R2_score: 0.6011612571432967  
Adjusted R2_score: 0.6008398290093625  
MSE: 134.883204825123  
MAE: 3.8520942387238177
```

Реализация случайного леса без подбора гиперпараметров



08 – Программная реализация

- Python3
- Jupyter Notebook
- Seaborn, Matplotlib
- Scikit-learn
- Statsmodel
- Pandas
- Numpy



<https://github.com/dkshi/aviation/>

Выводы

Что произошло:

- Статистический анализ выходной величины
- Сформирована задача
- Обработка признаков
- Корреляционный анализ
- Уравнение множественной линейной регрессии
- Улучшение модели

Что получено:

- Улучшаемая модель
- Прикладное использование
- Высокая актуальность



Спасибо за
внимание!