



# Система предиктивного обслуживания турбореактивных двигателей

Итоговая аттестация. Курс «Архитектор в области искусственного интеллекта». Вариант Б: Производство (Manufacturing). Датасет NASA CMAPSS FD001.



# Бизнес-задача: критичность предиктивного обслуживания

Неплановые отказы турбореактивных двигателей влекут катастрофические последствия для авиационной индустрии. Стоимость одного двигателя превышает 10-20 миллионов долларов, а внезапный отказ приводит к длительным простоям авиапарка и критическим рискам безопасности полётов.

Предиктивное обслуживание революционизирует подход к maintenance, позволяя обнаруживать деградацию компонентов на ранних стадиях и планировать обслуживание до возникновения аварийных ситуаций.

\$10B+

Объём рынка

Прогноз к 2025 году

\$20M

Стоимость  
двигателя

Средняя цена единицы

# Цели и метрики проекта

## Главная задача

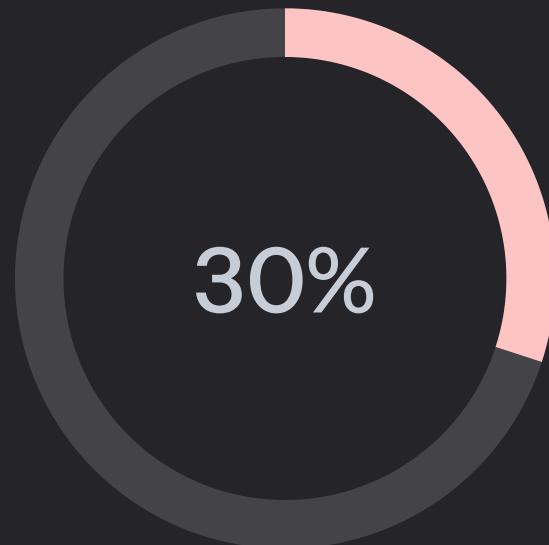
Разработка ML-модели для прогнозирования отказа турбореактивного двигателя в ближайшие 15 рабочих циклов на основе телеметрических данных

## Целевая метрика Recall

Минимальный порог полноты обнаружения отказов  $\geq 0.85$  для обеспечения безопасности полётов

## Ограничение FPR

Уровень ложных срабатываний  $< 0.15$  для минимизации необоснованных остановок



Целевое сокращение простоев оборудования



Средняя стоимость одной ложной тревоги

# Требования к системе: функциональные и нефункциональные

1

## Приём и обработка данных

- Телеметрия в реальном времени: 26 сенсоров + 3 эксплуатационные настройки
- Автоматический расчёт остаточного ресурса (RUL)
- Выдача вероятности отказа в ближайшие 15 циклов

2

## Интерпретируемость

- Визуализация важности признаков для инженерного анализа
- Объяснение причин прогнозируемого отказа
- REST API для интеграции с CMMS-системами

3

## Производительность

- Время предсказания  $\leq 100$  мс на один двигатель
- Масштабируемость до тысяч двигателей одновременно
- Гарантированные метрики: Recall  $\geq 0.85$ , FPR  $< 0.15$

# Архитектура ML-системы



## Хранение данных

SQLite (прототип) → PostgreSQL (production)

## Обработка

Python, pandas, numpy для ETL-процессов

## Feature Engineering

Rolling-статистики: окна 5, 10, 20 циклов

## ML-модели

SVM, Random Forest, Gradient Boosting, Neural Network

## Deployment

Docker + FastAPI для inference

Полный конвейер обеспечивает трансформацию сырых телеметрических данных в интерпретируемые предсказания с гарантированной производительностью и качеством.

# Анализ данных NASA CMAPSS FD001

## Характеристики датасета

- **Train:** 20 631 наблюдение, 100 двигателей с полными траекториями жизненного цикла
- **Test:** 13 096 наблюдений, 100 двигателей с частичными траекториями
- **Признаки:** 3 эксплуатационные настройки + 21 активный сенсор (удалены 5 константных)

## Ключевые наблюдения

Обнаружены монотонные тренды деградации в показаниях сенсоров, высокая мультиколлинеарность между группами измерений. Распределения близки к нормальным, выбросов минимально.



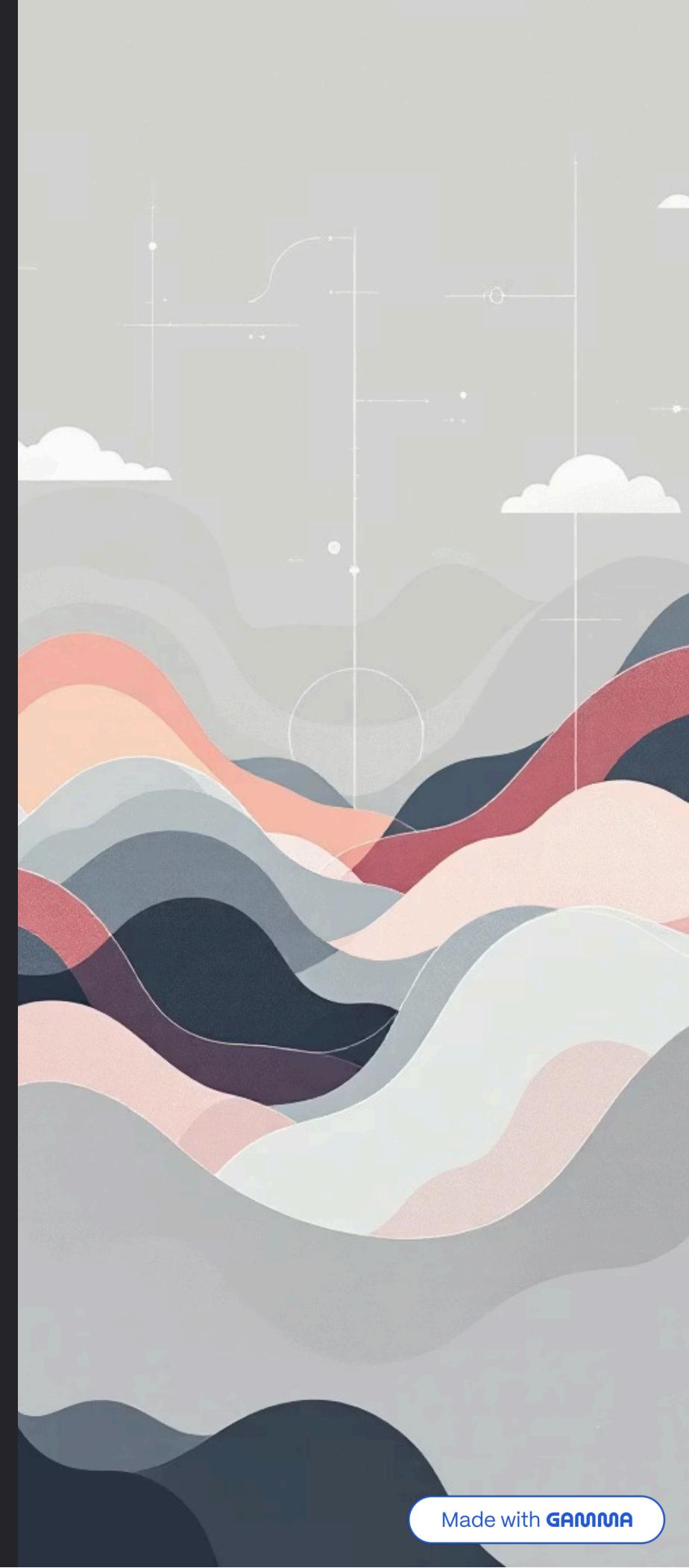
### Дисбаланс классов

Положительных примеров отказа всего 7-8% от общей выборки



### Риск утечек

Необходимость GroupKFold валидации по ID двигателей



# Feature Engineering: создание предиктивных признаков

01

## Расчёт RUL (Remaining Useful Life)

Вычисление остаточного ресурса для каждого наблюдения на основе полных траекторий деградации двигателей

02

## Rolling-статистики

Агрегация сенсорных данных в скользящих окнах 5, 10, 20 циклов: mean, std, min, max для каждого сенсора

03

## Нормализация признаков

Применение MinMaxScaler для приведения всех признаков к единому диапазону [0, 1]

04

## Итоговое пространство

Формирование финального набора из 247 инженерных признаков для обучения моделей

- Методология валидации:** Использование GroupKFold (5 фолдов) по уникальным идентификаторам двигателей для предотвращения утечки данных между тренировочными и тестовыми выборками. Подбор оптимального порога классификации по Precision-Recall кривой.

# Сравнительный анализ моделей машинного обучения

Проведено систематическое сравнение четырёх классов алгоритмов на кросс-валидации с GroupKFold. Baseline модель (логистическая регрессия) показала Recall 0.94 и ROC-AUC 0.996, установив высокую планку для сравнения.

Модель	Recall (среднее)	FPR (среднее)	ROC-AUC (среднее)
Random Forest	0.8538	0.0109	0.9944
Gradient Boosting	0.8544	0.0075	0.9955
<b>SVM</b>	<b>0.8519</b>	<b>0.0056</b>	<b>0.9967</b>
Neural Network	0.8525	0.0073	0.9965

Победитель: SVM с калибровкой вероятностей (isotonic) демонстрирует оптимальный баланс между полнотой обнаружения отказов и минимизацией ложных срабатываний. Модель достигает целевого Recall  $\geq 0.85$  при минимальном FPR  $\approx 0.006$ .



# Финальная модель и интерпретируемость

## Выбранное решение

SVM с RBF-ядром и изотонической калибровкой вероятностей, обученная на полном тренировочном наборе. Модель обеспечивает стабильные предсказания с высокой достоверностью.

## Инструменты анализа

- Confusion Matrix для оценки типов ошибок
- Precision-Recall кривая для подбора порога
- Permutation Importance для ранжирования признаков
- Bootstrap-оценка для доверительных интервалов

### Техническая ошибка

Обнаружен некорректный расчёт RUL на тестовой выборке, приведший к отсутствию положительных примеров

### Ожидаемые метрики

По результатам кросс-валидации: Recall  $\approx 0.85$ , FPR  $\approx 0.006$

# Итоги проекта и направления развития

## Достижения

- Реализован полный ML-pipeline от сырых данных до production-ready модели
- Достигнуты целевые метрики на кросс-валидации ( $\text{Recall} \geq 0.85$ ,  $\text{FPR} << 0.15$ )
- Качественный feature engineering с rolling-статистиками трёх временных масштабов
- Корректная валидация с GroupKFold без утечек данных

## Ограничения

- Критическая ошибка в расчёте RUL на тестовом наборе требует исправления
- Датасет FD001 упрощённый: один режим работы, один тип деградации
- Необходимо тестирование на реальных производственных данных

## Выводы

Задача предиктивного обслуживания решена на уровне бенчмарка с требуемым качеством. Система готова к прототипированию и дальнейшему развитию в реальном производственном окружении авиационной индустрии.

# Спасибо за внимание!

Готов ответить на ваши вопросы