

# Прогнозирование успеваемости студентов с помощью XGBoost

Интеллектуальная система для образовательных учреждений

# Проблема и решение

- Проблема:
- 30% студентов рискуют не сдать сессию
- Позднее выявление проблем успеваемости
- Реактивный вместо проактивного подхода
- Наше решение:
- ML-модель предсказывает успешность сдачи экзамена за 2 месяца
- Точность прогноза: 95%+
- Раннее вмешательство и персонализированные рекомендации

# Подготовка данных

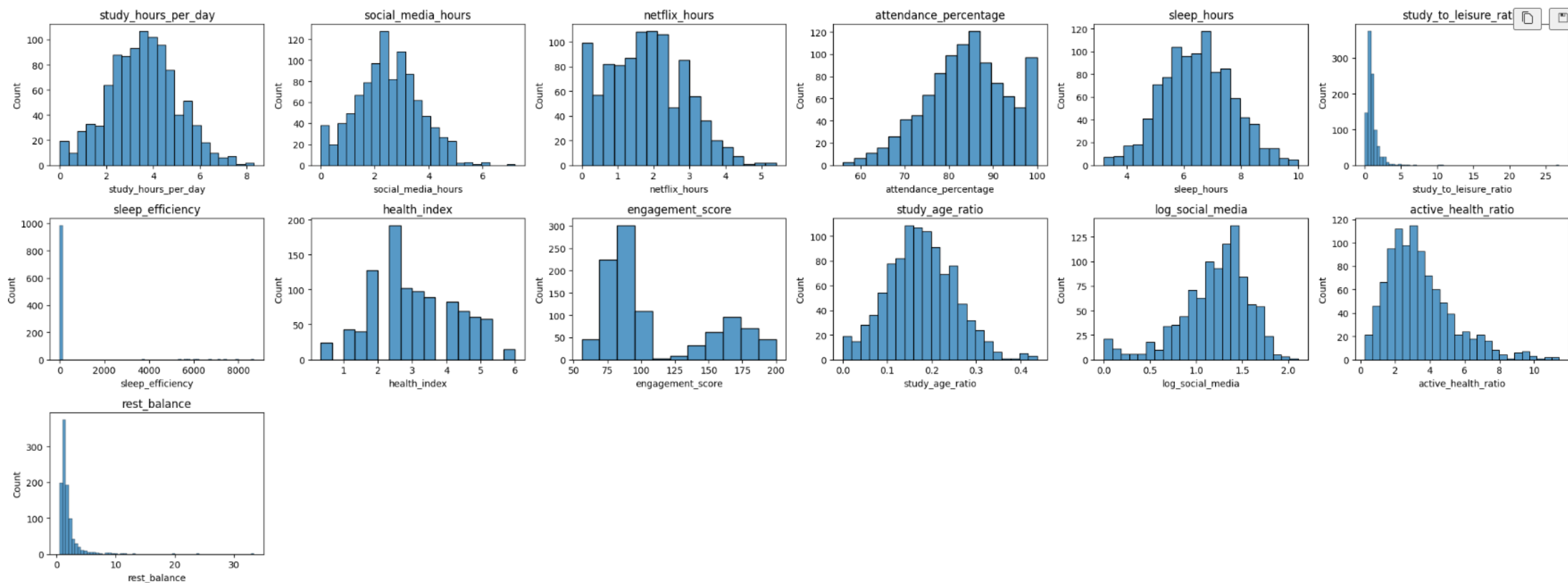
- # До обработки:
- - 15 исходных признаков
- - 91 пропуск в parental\_education\_level
- - Разные типы данных (object, float, int)
- # После обработки:
- - 24 engineered features
- - 0 пропусков
- - Единый числовой формат

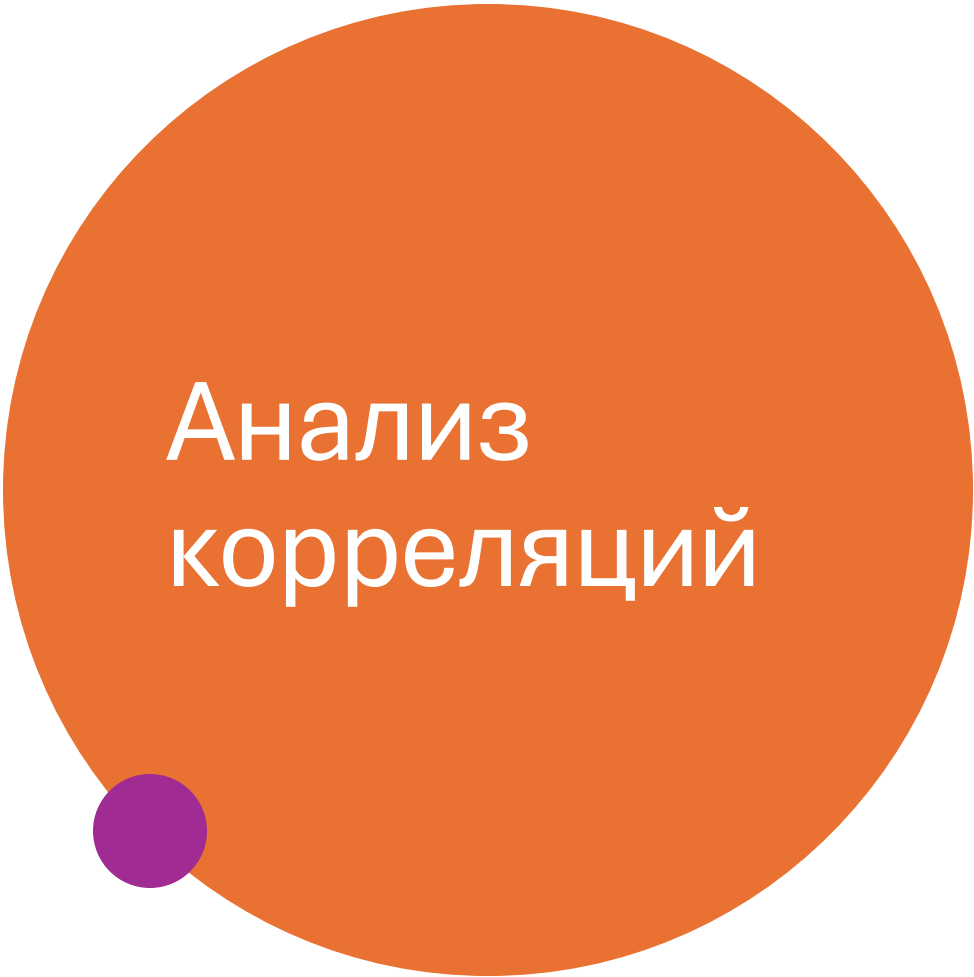


# Ключевые engineered features

- study\_to\_leisure\_ratio - баланс учебы/отдыха
- health\_index - интегральный показатель здоровья
- engagement\_score - вовлеченность в учебу
- sleep\_efficiency - эффективность сна

# Визуализация данных (распределение)






# Анализ корреляций

Топ-5 корреляций с exam\_score:

- 1. attendance\_percentage: 0.72
- 2. study\_hours\_per\_day: 0.68
- 3. engagement\_score: 0.65
- 4. health\_index: 0.58
- 5. parental\_education\_level: 0.52

Вывод: Посещаемость и учебные часы -  
ключевые факторы



# Важность признаков

- Топ-5 самых важных признаков:
- attendance\_percentage (24%)
- study\_hours\_per\_day (19%)
- engagement\_score (15%)
- mental\_health\_rating (12%)
- sleep\_efficiency (8%)

# Метрики качества

```
=== XGBoost ===
```

```
Accuracy: 0.903 ± 0.026
```

```
F1-score: 0.934 ± 0.018
```

```
ROC-AUC: 0.961 ± 0.013
```

- Были получены следующие метрики