Базовый минимум (TS + ML)

Основная идея:
аномалия = существенное отличие реального сетевого трафика от предсказанного моделью.

TS-подход (ARIMA / auto-ARIMA):

1. Построение временных рядов сетевого трафика (пакеты/байты/запросы в секунду).
2. ARIMA/ARIMAX прогноз активности.
3. Аномалии = большие отклонения факта от прогноза.

Базовый ML-классификатор:

1. Формирование признаков по окнам: объем трафика, резкие спайки, количество уникальных IP/портов, частоты событий.
2. ML-модель (Logistic Regression / RandomForest) для метки «норма / аномалия».
3. Метрики: F-score, Precision/Recall, confusion matrix.

Роскошный максимум (ML)

1. **Isolation Forest** по агрегированным признакам сетевого трафика.
2. Выделение редких паттернов (редкие IP, порты, протоколы).
3. Кластеризация аномальных сегментов трафика (TF-IDF событий + KMeans / DBSCAN).

Сервис

FastAPI backend:

1. загрузка данных сетевого трафика
2. базовая EDA (распределения, активности, частоты)
3. TS-прогноз ARIMA
4. детекция аномалий по разнице предсказанное vs фактическое
5. ML-классификатор для метки «норма / аномалия».

Streamlit UI:

1. графики нагрузок и аномалий
2. визуализация временных рядов
3. просмотр подозрительных участков трафика
4. (бонус) интерактивная кластеризация (PCA/t-SNE).

Output

1. список аномальных временных интервалов
2. графики прогноза ARIMA с отклонениями
3. визуализация сетевого трафика и кластеров.

Бонус

MLflow:

1. трекинг экспериментов
2. версии моделей (ARIMA, ML-классификатор).

Хранилище артефактов (S3):

1. модели (ARIMA, ML-классификатор)
2. словари шаблонов/событий
3. параметры препроцессинга
4. сохранённые EDA-отчёты.