

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА ПО КУРСУ “DATA SCIENCE”

Тема: Анализ новостных потоков для прогнозирования движения цены цифровых активов.

Слушатель: Новицкая Виктория

Разведочный анализ данных



Данные по цифровым активам (btcusdt_1m и ethusdt_1m):

- Метка времени (timestamp);
- Цена открытия (open_price);
- Максимальная цена (high_price);
- Минимальная цена (low_price);
- Цена закрытия (close_price);
- Объем торгов (volume).

Строк: 1017611

Данные по новостям (news):

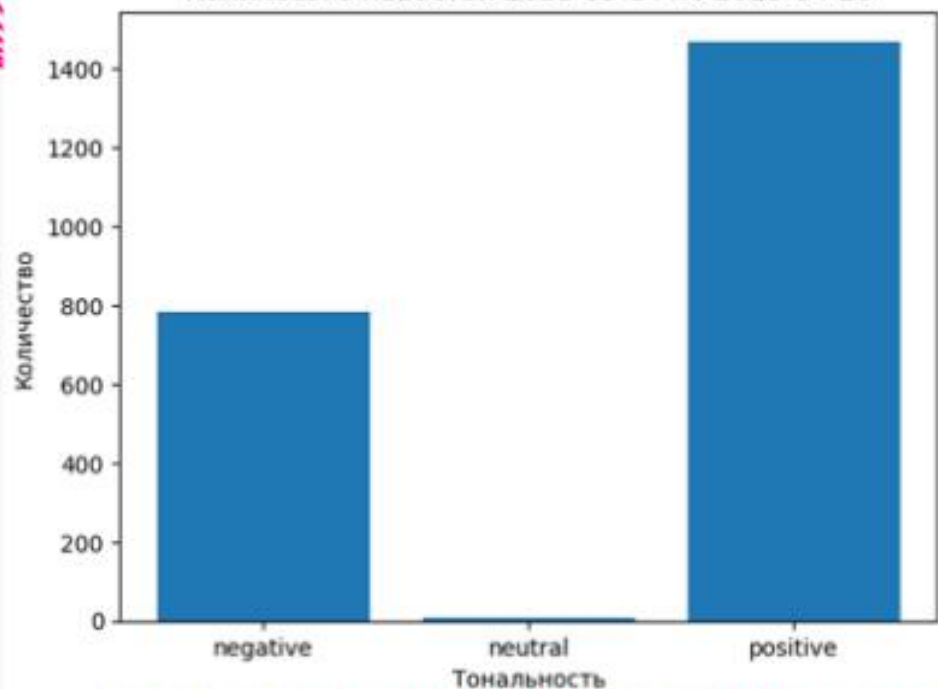
- Заголовок (Headline);
- Дата (Date);
- Текст новостных статей (Text).

Строк: 2466

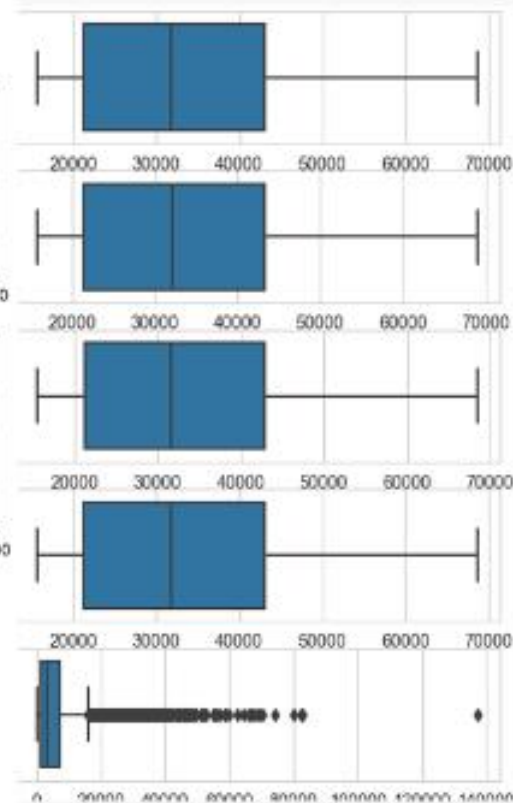
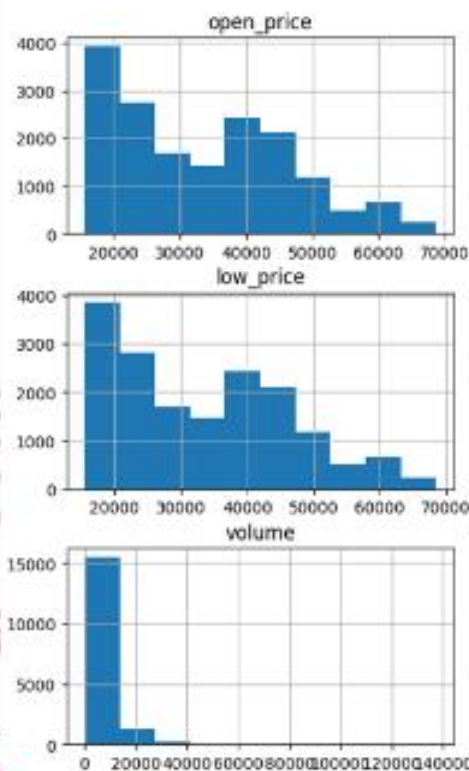
Данные за период с 14 мая 2021 года по 20 апреля 2023 года

Гистограммы распределения и диаграммы “ящик с усами”

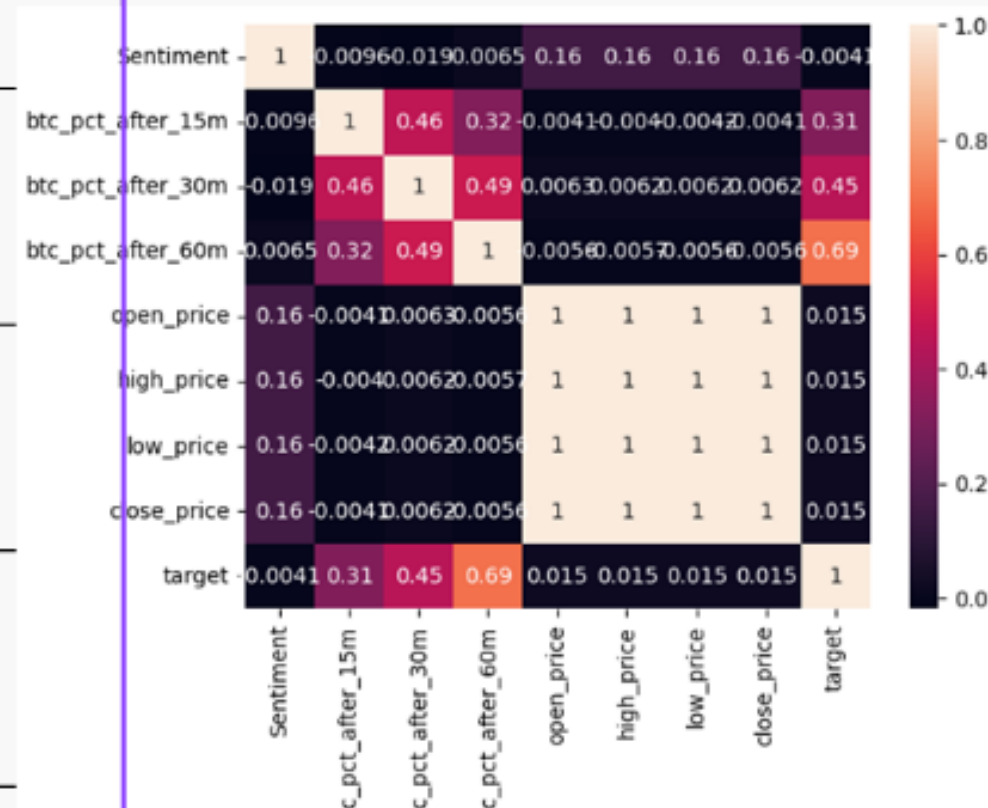
Количество новостей 2021-05-14 по 2023-04-20



Histogram of BTCUSDT



	precision	recall	F1-score	accuracy
Логистическая регрессия	0 - 71% 1 - 71%	0 - 73% 1 - 69%	0 - 72% 1 - 70%	71%
Дерево решений	0 - 63% 1 - 59%	0 - 58% 1 - 64%	0 - 60% 1 - 61%	61%
Случайный лес	0 - 67% 1 - 65%	0 - 66% 1 - 66%	0 - 67% 1 - 66%	66%
Градиентный бустинг	0 - 69% 1 - 69%	0 - 72% 1 - 66%	0 - 70% 1 - 67%	69%
Метод опорных векторов (SVM)	0 - 71% 1 - 69%	0 - 71% 1 - 69%	0 - 71% 1 - 69%	70%



Нейронные сети

MLP # 1

на обучающей выборке (71.83%) и на валидационной выборке (71.01%)

LSTM # 2

на обучающей выборке (71.26%) и на валидационной выборке (70.56%)

CNN # 3

на обучающей выборке (72.46%) и на валидационной выборке (70.27%)



Веб-приложение на Flask

← → ↻ ⓘ 127.0.0.1:5000

Enter the data to get a price movement prediction

Sentiment:

BTCUSDT After 15m:

BTCUSDT After 30m:

Open Price:

High Price:

Low Price:

Close Price:

← → ↻ ⓘ 127.0.0.1:5000/predict

```
{  
  "Prediction": "Sell"  
}
```



Перспективы развития

- Автоматизация анализа новостей
- Разработка стратегий торговли
- Факторный анализ
- Расширение диапазона прогнозируемых активов