

Определение влияния финансовых новостей на изменение цен Bitcoin-а

Сбертех 2024

Постановка задачи

Необходимость классификации тональности финансовых новостей и анализа их влияния на изменение цен Bitcoin для улучшения прогнозирования и оптимизации инвестиционных стратегий

Этапы:

- 1) Сбор данных
- 2) Предобработка текста
- 3) Построение модели
- 4) Оценка точности модели

Обзор литературы

BloombergGPT: A Large Language Model for Finance

- Решается не только целевая задача предсказания финансовых рядов.
- Используется архитектура Decoder.
- Для задачи классификации настроений в предложениях из финансовых новостей метрика 0.51 weight -F1

Enhanced news sentiment analysis using deep learning methods

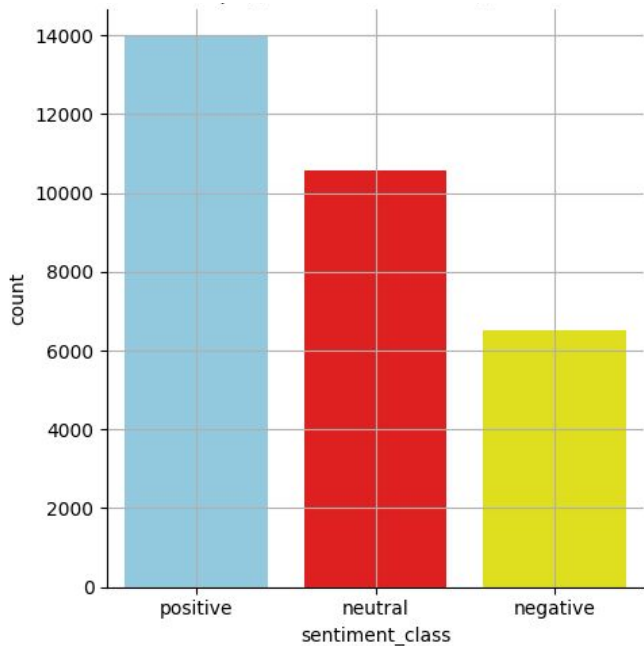
- Используется не SOTA подход к анализу текста.
- Актуальный для задачи подход к формированию датасета.
- Средняя точность прогноза получается порядка 0.76 accuracy

Описание Датасета

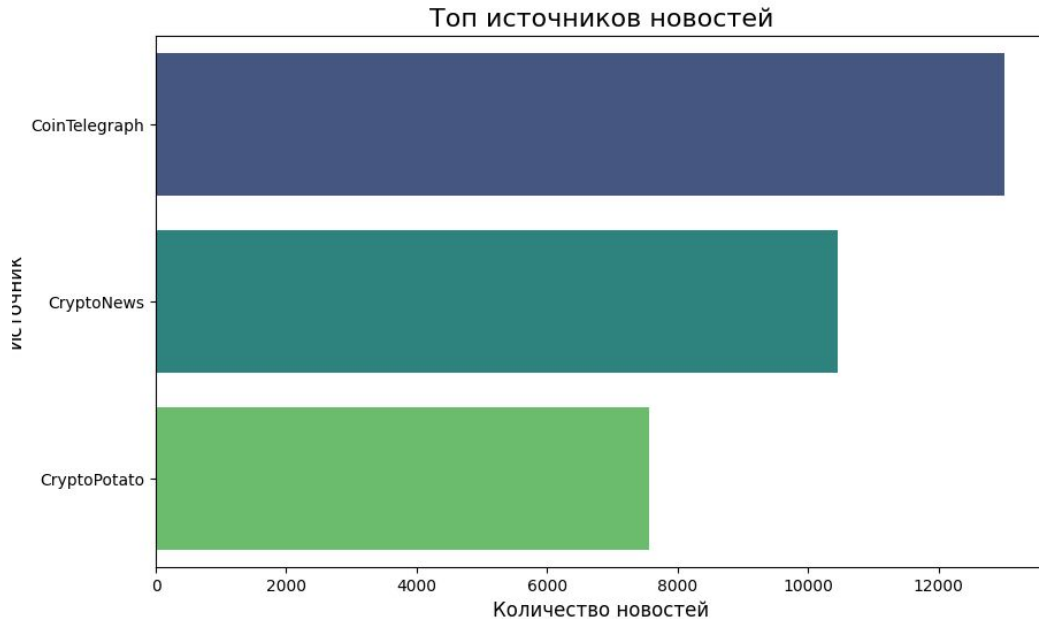
Данные о англоязычных новостях криптовалюты за более чем год в структурированном формате, включающем заголовки, текст, источник, тему и анализ настроений

date	sentiment	source	subject	text	title	url
2023-12-19 06:40:41	{'class': 'negative', 'polarity': -0.1, 'subjectivity': 0.6}	CryptoNews	altcoin	Grayscale CEO Michael Sonnenshein believes the SEC needs to approve spot Bitcoin exchange-traded fun...	Grayscale CEO Calls for Simultaneous Approval of Spot Products to Level the Field	https://cryptonews.comhttps://cryptonews.com/news/grayscale-ceo-calls-for-simultaneous-approval-of-s...
2023-12-19 06:03:24	{'class': 'neutral', 'polarity': 0.0, 'subjectivity': 0.0}	CryptoNews	blockchain	In an exclusive interview with CryptoNews, Manhar Garegrat, the Country Head for India and Global Pa...	Indian Government is Actively Collaborating With Crypto Industry: Liminal Custody's Country Head	https://cryptonews.comhttps://cryptonews.com/news/indian-government-is-actively-collaborating-with-c...
2023-12-19 05:55:14	{'class': 'positive', 'polarity': 0.05, 'subjectivity': 0.05}	CryptoNews	blockchain	According to the Federal Court ruling on December 18, former Binance CEO Changpeng 'CZ' Zhao has bee...	Judge Approves Settlement: Binance to Pay \$1.5 Billion to CFTC, CZ to Pay \$150 Million Fine	https://cryptonews.comhttps://cryptonews.com/news/judge-approves-settlement-binance-to-pay-1-5-billi...

Описание датасета



Распределение
классов в данных



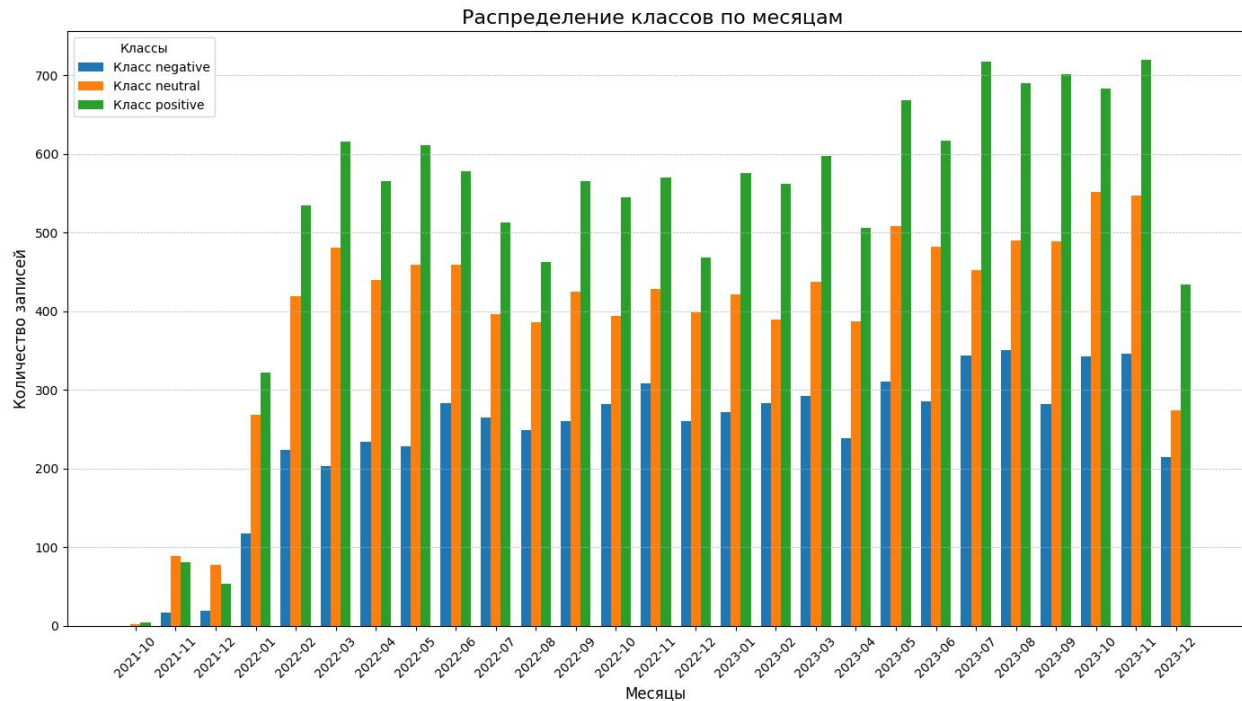
Распределение
новостей по
источникам

Описание датасета



Описание датасета

Баланс классов
сохраняется
независимо от месяца



TF-IDF + autoML

Параметры TF-IDF

5000 N-grams

Unigram

Bigram

Параметры autoML

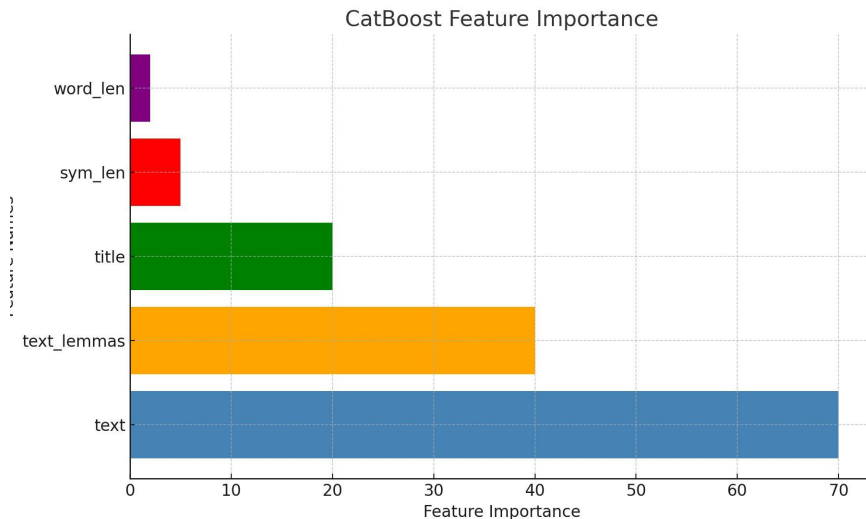
Параметр	Значение
colsample_bytree	0.6649
learning_rate	0.1740
max_bin	255
min_child_samples	3
n_estimators	159
num_leaves	19
reg_alpha	0.00098
reg_lambda	0.00676

CatBoost

Обучение проходило на таких фичах как:

- Длина новости в словах
- Длина новости в символах
- Лемматизированный текст новости и заголовка
- Необработанный текст новости

Метрика	Значение
F1 (macro average)	0.8465
Accuracy	0.8589



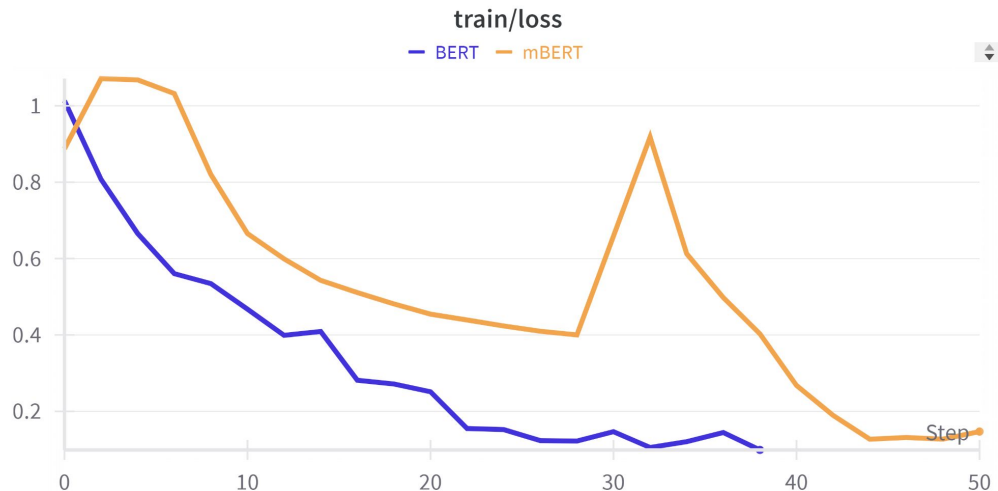
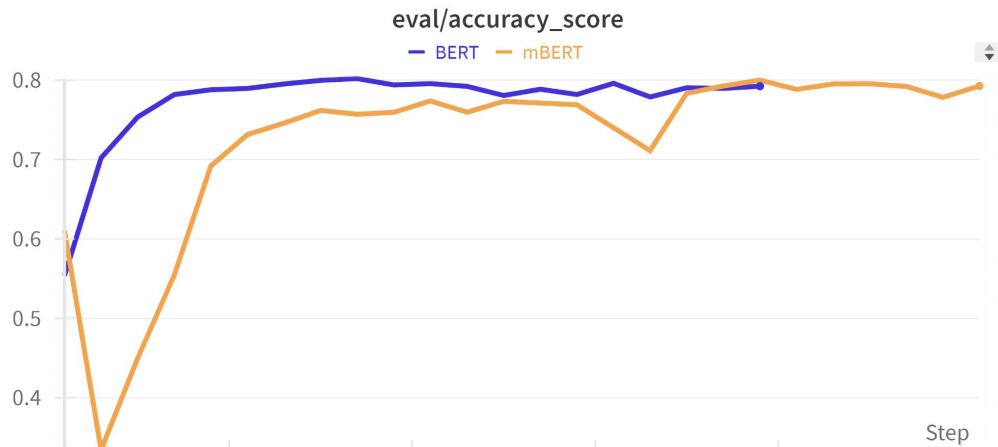
BERT

train - 70% датасета

test - остальные 30%

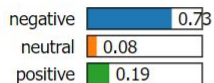
Обучение на
лемматизированной паре
заголовок - новость

Модель	Accuracy	F1-Score
BERT	0.7989	0.7989
mBERT	0.7733	0.7611



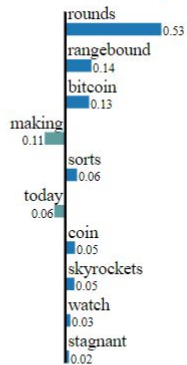
Интерпретируемость

Prediction probabilities



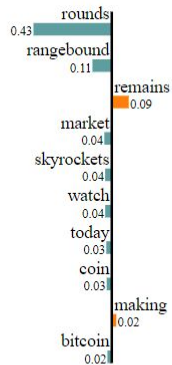
NOT negative

negative



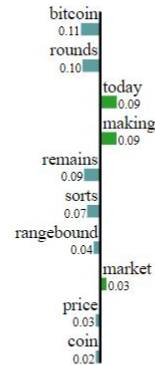
NOT neutral

neutral



NOT positive

positive

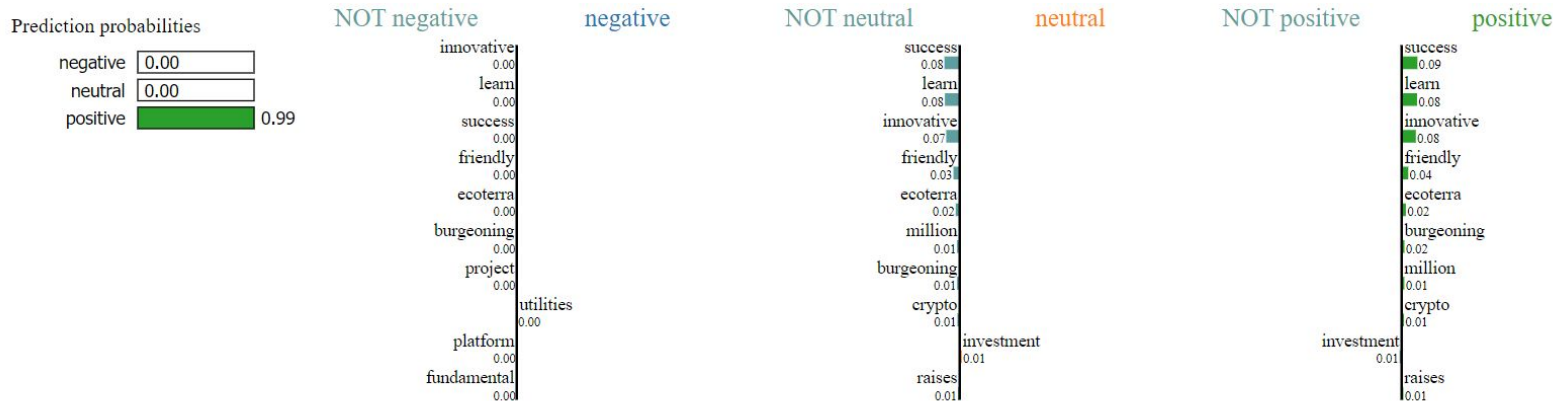


Text with highlighted words

bitcoin stagnant skyrockets market watch bitcoin price remains rangebound coin making sorts rounds today

Bitcoin's price remains rangebound but there's one coin that's making all sorts of rounds today

Интерпретируемость



Text with highlighted words

friendly crypto startup raises million ecoterra friendly crypto project raised million presale investment innovative platform burgeoning success fueled fundamental utilities learn

Ecoterra, an eco-friendly crypto project, has raised over \$5.9 million in presale investment. This innovative platform's burgeoning success is fueled by its fundamental utilities. Learn more here.

Результаты

Модель	Accuracy	F1-Score
CatBoost	0.8589	0.8465
BERT	0.7989	0.7989
TF-IDF + AutoML	-	0.79
BloombergGPT	-	0.51
RNN + LSTM	0.76	-

Команда



Березин Даниил



Кривенко Михаил