

Предзащита дипломной работы

Иванов Григорий M02-030

Тема работы

Построение моделей краткосрочного предсказания цены на криптовалюты
методом XGBoost

Введение

- Рынок криптоактивов является очень привлекательным для бизнеса
- Выявление трендов рынка поможет успешно вести бизнес
- Краткосрочное предсказание цены криптовалют является частью поиска закономерности крипторынка

XGBoost

- XGBoost - одна из самых популярных реализаций алгоритма градиентного бустинга деревьев решений
- Основания выбора XGBoost: доступность, популярность, актуальность

Начальные данные

- Получены поминутные данные по 14 популярным криптовалютам
- Для каждой криптовалюты имеются следующие значения:
 - Count
 - Open
 - High
 - Low
 - Close
 - Volume
 - VWAP

Lag features

$j = 0, 1, 2, \dots, 15 \text{ min}$

$$1. \frac{Close(t-j)}{Close(t)}$$

$$2. \frac{Close(t-j) - Open(t-j)}{Close(t)}$$

$$3. \frac{High(t-j) - Low(t-j)}{Close(t)}$$

$$4. \frac{Volume(t-j)}{Volume(t)}$$

$$5. \frac{VWAP(t-j)}{VWAP(t)}$$

$$6. \frac{VWAP(t-j)}{Close(t)}$$

Window averages

window = 1h, 3h, 6h, 9h, 12h ,24h

$$1. \frac{\text{average}(\text{VWAP}(\text{window}))}{\text{Close}(t)}$$

$$2. \frac{\text{average}(\text{Count}(\text{window}))}{\text{Count}(t)}$$

$$3. \frac{\text{average}(\text{Volume}(\text{window}))}{\text{Volume}(t)}$$

$$4. \frac{\text{min}(\text{Close}(\text{window}))}{\text{Close}(t)}$$

$$5. \frac{\text{max}(\text{Close}(\text{window}))}{\text{Close}(t)}$$

$$6. \frac{\text{stddev}(\text{VWAP}(\text{window}))}{\text{Close}(t)}$$

$$7. \frac{\text{stddev}(\text{Count}(\text{window}))}{\text{Count}(t)}$$

$$8. \frac{\text{stddev}(\text{Volume}(\text{window}))}{\text{Volume}(t)}$$

$$9. \frac{\text{stddev}(\text{Close}(\text{window}))}{\text{Close}(t)}$$

Разбиение данных и target

- Данные разбиты на train, validation и test датасеты в пропорциях 60%, 20%, 20%

- Target: $\log\left(\frac{Close(t + 16)}{Close(t + 1)}\right)$

XGBoost с стандартными параметрами

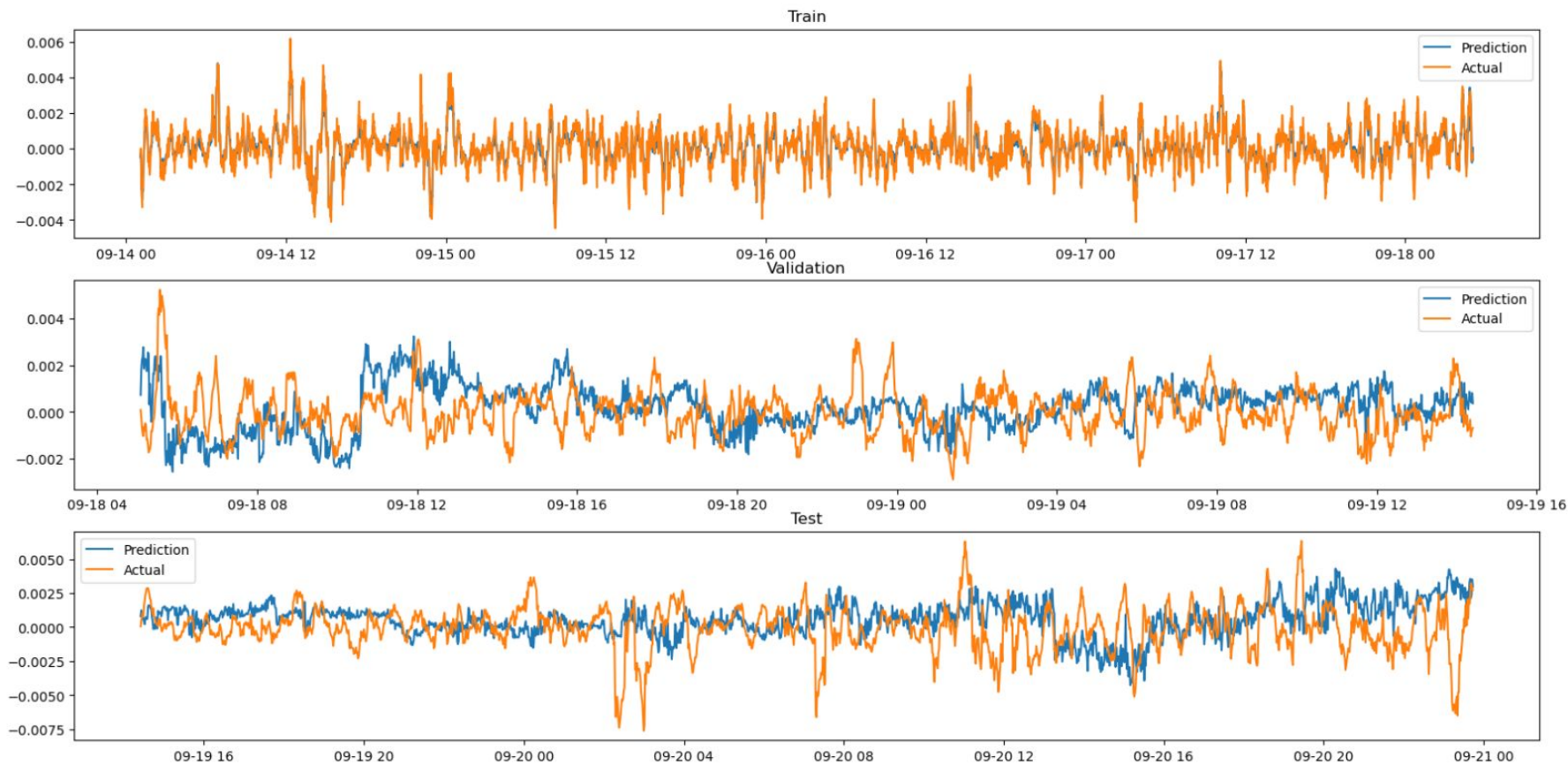
- В качестве метрики оценки модели выбран RMSE

```
*==Default model results==*
```

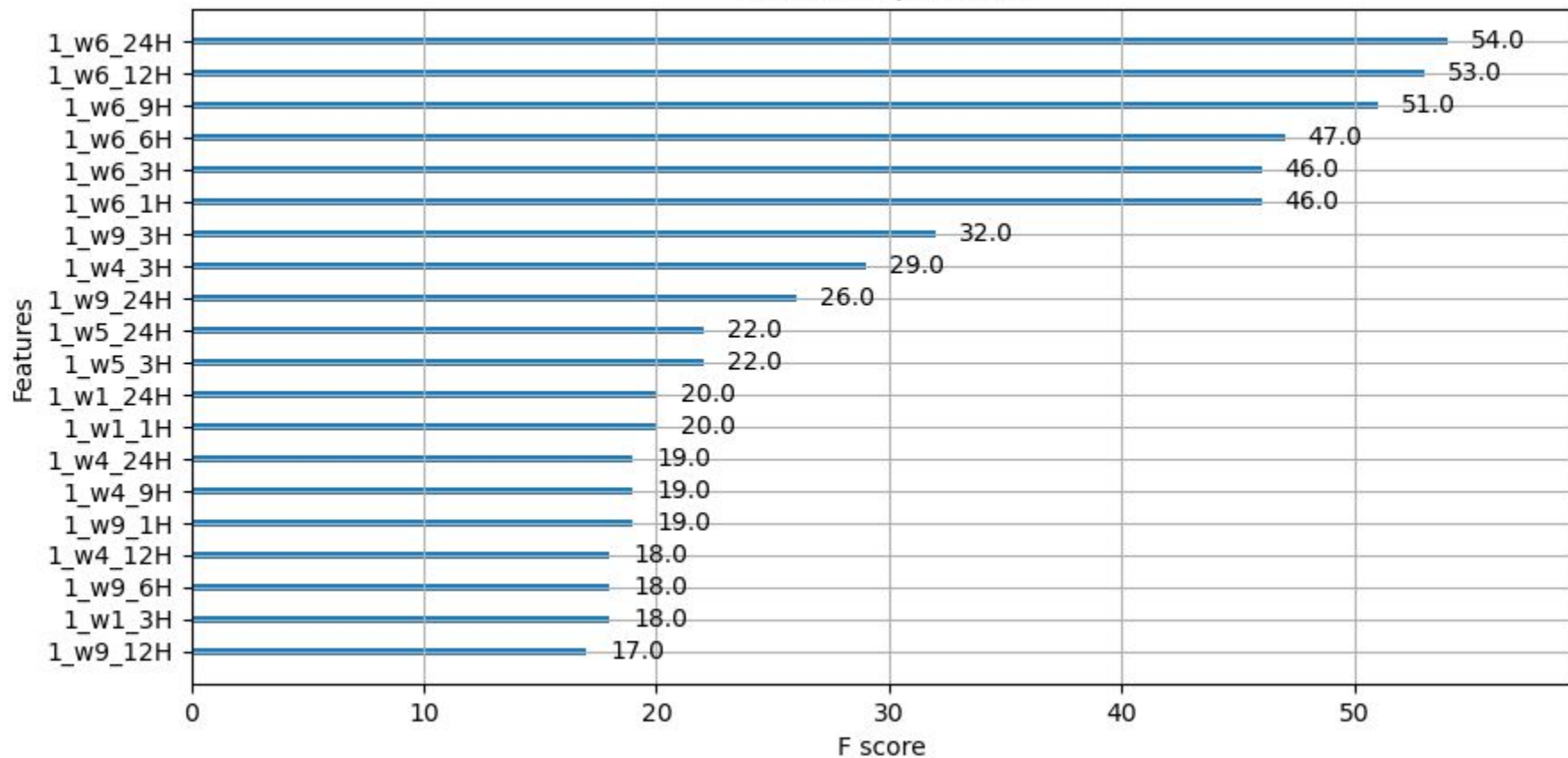
```
validation score: 0.0013252785897935632
```

```
test score: 0.002166911981285623
```

```
Execution time: 0:00:02
```



Feature importance



Тюнинг XGBoost через RandomSearch

1. Обучение на train
2. Поиск наилучшей модели на validation
3. Обучение на train + validation для проверки на test

Сложности

- Поиск решений удовлетворительных по времени и вычислительной мощности
- Подбор параметров для RandomSearchCV

Дальнейшие действия

- Визуализация результатов предсказания Target
- Тюнинг параметров через RandomSearch(возможно через GridSearch)
- Предсказания target криптовалюты по features другой
- Добавление новых features в датасет

Предварительные результаты и выводы

- В датасет требуется добавить более широкий диапазон для moving averages. Также необходимо добавить отдельные features связанные с VWAP и стандартным отклонением других параметров
- На стандартных параметрах модель обучается с ошибкой порядка не больше Target
- При должном тюнинге модели можно добиться существенного прогресса