**УДК 519.246'8**

# ПРОГНОЗИРОВАНИЕ КУРСА КРИПТОВАЛЮТ С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

**Скударь И.М.** (Национальный исследовательский университет ИТМО)
**Научный руководитель – ктн, доцент Маркина Т.А.**
(Национальный исследовательский университет ИТМО)

Работа направлена на разработку приложения по прогнозированию курса криптовалют. Данная работа основана на использовании алгоритмов машинного обучения.

**Введение.**

Любой рынок будь то крипто-рынок, биржа облигаций, акций и так далее подвержены изменению. Данные изменения диктуются в большинстве случаев самим рынком, то есть активностью пользователей. Немаловажным фактом является новостной фон, который стимулирует рост либо спад рынка. К примеру, вспомните твиты Илона Маска, которые способствуют масштабным изменениям рынка (например, монетка DOGE).

**Основная часть.**

Можно выделить основные задачи на данный проект:

- сбор исторических данных (парсинг и сбор релевантных новостей)
- построение моделей прогнозирования временных рядов и классификации текста
- вывод результатов в telegram бота

На торговлю на крипто рынке можно смотреть с разных сторон. Большим преимуществом у трейдера было бы знание новостей, связанных с рынком. Новостной фон зачастую очень сильно коррелирует с изменчивостью цены активов, а также влияет на поведение рынка в целом. Хочется дать трейдерам возможность наблюдать на рынок с разных сторон – графики, цифры и новости.

Сбор исторических данных производится с помощью api yahoo finance. Там есть все необходимые данные: временные ряды курса определенной валюты и новости.

Для реализации классификации новостей используется модель, основанная на совокупности TF-IDF метрики и логистической регрессии. TF-IDF (TF — term frequency, IDF — inverse document frequency) — статистическая мера, используемая для оценки важности слова в контексте документа, являющегося частью коллекции документов. Вес некоторого слова пропорционален частоте употребления этого слова в документе и обратно пропорционален частоте употребления слова во всех документах коллекции.

Мера TF-IDF часто используется в задачах анализа текстов и информационного поиска, например, как один из критериев релевантности документа поисковому запросу, при расчёте меры близости документов при кластеризации. TF (term frequency — частота слова) — отношение числа вхождений некоторого слова к общему числу слов документа. Таким образом, оценивается важность слова t_i в пределах отдельного документа. IDF (inverse document frequency — обратная частота документа) — инверсия частоты, с которой некоторое слово встречается в документах коллекции. Учёт IDF уменьшает вес широкоупотребительных слов. Для каждого уникального слова в пределах конкретной коллекции документов существует только одно значение IDF.

Данная модель была обучена на ~4500 заголовков новостей, классифицированных вручную на позитивные, нейтральные и негативные. Оценка модели производилась метрикой Accuracy.

**Выводы.**

Результаты данного проекта можно использовать для торговли криптовалютой новичкам, которые не хотят углубляться в поиск новостей и анализ рынка в целом. Данное решение

можно масштабировать на рынок акций, облигаций, т.е. оно не заточено конкретно под криптовалюту.

| | |
|---|---|
| Скударь И.М. (автор) | Подпись |
| Маркина Т.А. (научный руководитель) | Подпись |