

Санкт-Петербургский государственный университет

*Филиппов Марк Дмитриевич*

Выпускная квалификационная работа

Применение нейронных сетей для  
прогнозирования цен на рынке  
криптовалют

Уровень образования: бакалавриат

Направление *02.03.03 «Математическое обеспечение и администрирование  
информационных систем»*

Основная образовательная программа *СВ.5006.2017 «Математическое обеспечение и  
администрирование информационных систем»*

Научный руководитель:  
к.ф.-м.н., доцент кафедры информатики, Григорьев Д. А.

Рецензент:  
д. э. н., профессор Прохоров А. Б.

Санкт-Петербург  
2021

Saint Petersburg State University

*Mark Philippov*

Bachelor's Thesis

# Application of neural networks for cryptocurrencies price prediction

Education level: bachelor

Speciality *02.03.03 "Software and Administration of Information Systems"*

Programme *CB.5006.2017 "Software and Administration of Information Systems"*

Scientific supervisor:  
C.Sc., docent. Dmitriy Grigoriev

Reviewer:  
Ph.D., prof. Artem Prohorov

Saint Petersburg  
2021

# Оглавление

<b>1. Введение</b>	<b>4</b>
<b>2. Постановка задачи</b>	<b>5</b>
2.1. Вопросы исследования . . . . .	5
<b>3. Обзор</b>	<b>7</b>
<b>4. Данные</b>	<b>11</b>
4.1. Технические индикаторы . . . . .	11
4.2. Фундаментальные показатели . . . . .	14
4.3. Макроэкономические показатели . . . . .	17
<b>5. Архитектуры нейронных сетей</b>	<b>18</b>
5.1. Multi-layer perceptron . . . . .	18
5.2. LSTM[1] . . . . .	18
5.3. GRU[2] . . . . .	21
5.4. biLSTM[3] . . . . .	22
5.5. sLSTM . . . . .	22
<b>6. Методология</b>	<b>24</b>
6.1. Подготовка данных . . . . .	24
6.2. Подготовка и запуск моделей . . . . .	29
<b>7. Результаты</b>	<b>31</b>
7.1. Обсуждение результатов . . . . .	31
<b>8. Заключение</b>	<b>33</b>
<b>Список литературы</b>	<b>34</b>
<b>Приложение</b>	<b>38</b>

# 1. Введение

В течение последних нескольких лет в мире отмечается устойчивый тренд на так называемые криптовалюты. Начало этому положила первая в мире криптовалюта - Bitcoin [4]. Если в начале крупные игроки не интересовались этим видом активов, то сейчас криптовалюты заняли свою нишу и многие участники торгов добавляют их в свои портфели. И естественным образом перед ними встает вопрос о прогнозировании цены на них.

В данный момент уже существует более чем 9000 разных криптовалют [5], а общая капитализация рынка превышает 1.9 трлн долларов США. Все это многообразие криптовалют обладает разными характеристиками (далее рассмотрим какими именно), но они едины в основополагающих принципах анонимности и децентрализованного устройства сети.

Предсказывание финансовых рынков - очень сложная задача. Прежде всего котировки ценных бумаг (и криптовалют, в частности) отличаются большим количеством шумов и наличием сложных внутренних взаимосвязей. Из-за этого стандартные статистические способы предсказания доходности не дают желаемого результата [6]. Поэтому в работе будем использовать нейронные сети, которые на данный момент являются state of the art в области предсказания финансовых временных рядов.

## 2. Постановка задачи

Целью работы является:

1. Собрать широкий набор данных по криптовалютам на основе технических, фундаментальных и макроэкономических показателей, а затем с помощью статистических методов определить узкий набор наиболее важных для прогнозирования компонентов.
2. Сделать обзор современных архитектур нейронных сетей, реализовать их для задачи прогнозирования доходности криптовалют и модернизировать их с помощью кросс-валидации и добавления большего числа признаков.
3. Определить на собранных данных архитектуру нейронной сети, которая наиболее качественно решает задачу прогнозирования доходности криптовалют.

### 2.1. Вопросы исследования

Во многих работах в области предсказания доходностей криптовалют есть ряд недостатков:

1. Работы [7][8][9][10] рассматривают только одну криптовалюту Bitcoin или Ethereum, и не обобщают свой результат на другие активы.
2. Также в работах [7][8][9][11] не была использована кросс-валидация, что уменьшает точность итоговых результатов.
3. В работе [11] в качестве входных признаков были использованы технические индикаторы, но не было применено алгоритма для уменьшения размерности датасета, что могло привести к тому, что бесполезные признаки замедляли работу модели, при этом почти не улучшая ее.
4. Также только в одной работе [8] были рассмотрены фундаментальные характеристики криптовалюты в качестве входных дан-

ных, в остальных же статьях обычно использовались либо технические индикаторы, либо же OHLCV (Цена открытия, наивысшая цена, наименьшая цена, цена закрытия, объём торгов).

Исходя из этого формулирую основные вопросы исследования:

1. Все ли технические индикаторы одинаково влияют на итоговое предсказание модели, а если не все, то какие влияют больше всего?
2. Улучшаться ли результаты исходных моделей при добавлении разнородных фундаментальных и макроэкономических признаков к техническим индикаторам?
3. Какие точные результаты работы моделей, после усреднения по всем криптовалютам и применения кросс-валидации?

### 3. Обзор

- Neural networks in finance and economics forecasting [12]

В этой статье рассмотрены основные принципы использования нейронных сетей для предсказания финансовых рядов, на примере акций. Их эффективность объясняется способностью нейронных сетей улавливать нелинейные зависимости, что как раз нужно в финансовой сфере, так как модели, которая строго описывает будущее поведение финансовых активов по каким-либо параметрам, не существует. Отмечена нелинейная зависимость между ценами на акции, их дивидендами и объемами торгов, а также нелинейная зависимость между индексом акций и макроэкономическими индикаторами.

- ETH analysis and predictions utilizing deep learning [7]

Применение рекуррентных нейронных сетей для предсказания цены на криптовалюту Ethereum и последующее сравнение результатов. В качестве входных данных были использованы котировки закрытия торговой сессии и средневзвешенные котировки на протяжении дня. Сравнивались модели: CNN-2l, CNN-3l, LSTM, sLSTM, BiLSTM, GRU, лучшие результаты на тестирующей части датасета показала обычная LSTM сеть. Архитектурно она из себя представляла входной LSTM слой из 50 нейронов с функцией активации - гиперболический тангенс и вероятностью "выключения нейронов" - 0.2. Также была применена L2 регуляризация со значением 0.0001 для решения проблемы переобучения.

- Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators [11]

В этой работе авторы использовали нейронные сети для предсказания дальнейшего тренда движения криптовалют. Из особенностей стоит отметить высокочастотность предсказаний (использовались минутные котировки), также была использована не одна

валюта, а шестерка самых популярных. В качестве входных данных были выбраны разнообразные технические индикаторы, всего 18 штук. Были использованы всего 4 архитектуры нейронных сетей: CNN, CLSTM, MLP, RFBNN. Наилучший результат показала CLSTM сеть, причем абсолютно на всех валютах, что подтверждает предположение об эффективности сетей с памятью. Также даны конкретные настройки сети со всеми параметрами.

- Forecasting the equity risk premium: The role of technical indicators [13]

Большая работа, в которой была проверена предсказательная способность технических индикаторов. Хотя в научных работах, до недавнего времени, они встречались достаточно редко и в основном использовались практиками биржевой торговли. Несмотря на это, результаты работы заключаются в том, что технические индикаторы можно использовать для предсказания доходностей активов самостоятельно, так и вместе с макроэкономическими величинами. Причем совместное использование этих двух способов дает наилучший результат, так они улавливают различные виды зависимостей и не взаимозаменяемы друг другом.

- On stock return prediction with LSTM networks [14]

В этой работе в центре внимания была LSTM сеть, исследовалось два подхода: регрессионный и классификационный (можно предсказывать не точно значение доходности на следующий временной период, а направление изменения). При этом в качестве данных были использованы индексы акций США, Бразилии и Швеции. Для нас представляет интерес, как статья, где ни одна архитектура LSTM (тривиальная, глубокая, с разными функциями активации) не смогла предсказать индекс S&P500 и индекс BOVESPA на уровне превышающем статистическую погрешность, хотя шведский индекс OMX все же удалось предсказать с более высокой точностью. Автор объясняет это эффективностью рынка первых



двух стран.

- A Gated Recurrent Unit Approach to Bitcoin Price Prediction [8]

В этой статье авторы проверяют предсказательную способность нейронной сети GRU и сравнивают ее с LSTM (GRU - более современная модель, в которой меньше входных фильтров и как следствие меньше производимых операций, но, естественно, это влияет на способность улавливать скрытые зависимости). В качестве входных данных были выбраны 20 различных признаков, большая часть из которых носят более фундаментальный характер. При этом, это одна из немногих работ, где был поднят вопрос мультиколлинеарности, ("похожести" значения признаков). Для каждого из признаков был подсчитан VIF - variance inflation factor, и если значение VIF было выше 10, то признак отбраковывался, что помогло сократить датасет на 25%, и, как следствие, сильно повысить скорость обучения нейронной сети. Также в этой работе есть обоснование выбора различных архитектурных решений, чего не наблюдается у других авторов. В результате проделанной работы получилось, что наилучшие результаты показала нейронная сеть с архитектурой GRU. Причем "тривиальная" сеть из 1-го слоя GRU и 1-го Dense слоя имеет ошибку меньше чем архитектура со скрытым слоем GRU

- Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators [15]

В этой статье авторы строят нейронные сети, которые выполняют задачи классификации: предсказывается направление, куда пойдет цена. При этом используется очень интересная архитектура. Изначальный набор признаков был разбит на две группы: технические индикаторы и макроэкономические параметры. По каждой группе обучалась отдельная нейронная LSTM сеть, а уже потом при согласованной работе этих сетей, делается предсказание на направление движения цены. Эта гибридная LSTM сеть показала

результаты лучше, чем простая LSTM сеть построенная по полному набору признаков.

## 4. Данные

Рынок криптовалют - очень молод, при этом в самом начале его становления, количество криптовалют было невелико. Так как в основе любого машинного обучения должен лежать обучающий и проверяющий датасет (train и test), выберем несколько криптовалют для которых и будем проводить все дальнейшие исследования, главный критерий - большое количество исторических данных. Нашим условиям удовлетворяют: Bitcoin (BTC), Ripple(XRP), Dash(DASH), Monero(XMR), Litecoin(LTC).

В основе работы любой нейронной сети лежат данные. А выбор этих данных и их подготовка, не менее важны, чем сама предсказательная часть, так как мусор на входе всегда даст мусор на выходе. После анализа литературы на тему использования нейронных сетей в предсказании доходности криптовалют, было принято решение совместить несколько типов данных для более точной работы алгоритма. Поэтому итоговый датасет будет состоять из трех частей:

- Технические индикаторы
- Фундаментальные показатели
- Макроэкономические показатели

Далее поговорим и рассмотрим каждый пункт по отдельности.

### 4.1. Технические индикаторы

Технический индикатор - функция, которая построена по прошлым значениям торгов. Техническими индикаторами занимается технический анализ - один из способов анализировать финансовые временные ряды. Но в чистом виде, на данный момент, применяется довольно редко, так как финансовые рынки характеризуются большим количеством шумов и нелинейных зависимостей [16]. Но есть ряд работ, где технические индикаторы используются в качестве входных данных в нейронных сетях [11][17][18], при этом сети с техническими индикаторами и OHLCV

данными (Open, High, Low, Close, Volumes) показывают в итоге ошибку меньшую чем сети просто с данными о предыдущих доходностях [19]. При этом именно нейронные сети помогут уловить (конечно же если они существуют) нелинейные зависимости между техническими индикаторами. Все они, в зависимости от своего принципа работы и исходных данных, делятся на несколько классов:

- Индикаторы моментума (скорости рынка)
- Индикаторы объема
- Индикаторы волатильности
- Индикаторы тренда

С помощью библиотеки [20] рассчитаем технические индикаторы для наших криптовалютных активов. Будем использовать некоторые основные индикаторы из каждого класса.

Class	Indicator	Parameters
Momentum	RSI	window = 14
	TSI	window slow = 25, window fast = 13
	ROC	window = 12
Volume	VPT	-
	EoM	window=14
	OBV	-
	ADI	window=14
Volatility	ATR	window=14
	Bollinger High Band Indicator	window=20, window dev=2
	Bollinger Low Band Indicator	window=20, window dev=2
Trend	EMA	window = 14
	MACD	window slow=26, window fast=12, window sign=9
	Parabolic SAR	step=0.02, max step=0.2
	WMA	window=5, window=10, window=20, window=30, window=60
	SMA	window=5, window=10, window=20, window=30, window=60

Таблица 1: Используемые технические индикаторы

## 4.2. Фундаментальные показатели

Фундаментальный анализ - метод измерения внутренней стоимости ценных бумаг, путем изучения связанных экономических и финансовых факторов [21]. Важной частью фундаментального анализа является изучение финансовых отчетов компании эмитента для анализа: уровня чистой прибыли, выручки, долговой нагрузки и многих других показателей компании. При этом эти показатели успешно используются в предсказывании доходности акций [22]. Так как криптовалюта - не является долей в компании и фактически не имеет никакой "внутренней" стоимости, можно анализировать только разнообразные показатели криптовалютной сети, которые также как технические и макроэкономические индикаторы, постоянно изменяются. При этом эти показатели - очень специфичны, и применяются только к криптовалютам, большая часть из них относится к внутреннему устройству (количество майнеров в сети, сложность вычисления блока, суммарные транзакции в сети и тд.). Из-за того, что рассматриваем не одну криптовалюту, а несколько, то в силу разной архитектуры и внутреннего устройства, для некоторых криптовалют невозможно подсчитать все показатели. Для Monero, в силу повышенных требований к приватности этой валюты, будет невозможно подсчитать целый ряд метрик, связанных с количеством транзакций, их объемом (речь идет именно о суммарном объеме переводов в криптовалютной сети, не путать с объемом торгов) и тп. Таблица 2 была составлена по данным с сайта [23], используя общедоступные данные

Feature	Description
AdrActCnt	The sum count of unique addresses that were active in the network that day.
AssetEODCompletionTime	Last fully calculated EOD metrics time
BlkCnt	The sum count of blocks created

	that day that were included in the main chain.
BlkSizeByte	Block size in bytes
BlkSizeMeanByte	The mean size of all blocks created that day.
CapMVRVCur	The ratio of the sum USD value of the current supply to the sum "realized" USD value of the current supply.
CapMrktCurUSD	The sum USD value of the current supply. Also referred to as network value or market capitalization.
CapRealUSD	The sum USD value based on the USD closing price on the day that a native unit last moved for all native units.
DiffMean	The mean difficulty of finding a hash that meets the protocol-designated requirement that day.
FeeMeanNtv	The mean fee per transaction in native units that day.
FeeMeanUSD	The USD value of the mean fee per transaction that day.
FeeMedNtv	The median fee per transaction in native units that day.
FeeMedUSD	The USD value of the median fee per transaction that day.
FeeTotNtv	The sum of all fees paid to miners, transaction validators, stakers and/or block producers that day.
HashRate	The mean rate at which miners are solving hashes that day.
IssContNtv	The sum of USD value issued that day.
IssContPctAnn	The percentage of new native units issued on that day, extrapolated to one year (i.e., multiplied by 365), and divided by the current supply on that day
IssContUSD	The sum USD value of new native units issued that day
IssTotNtv	The sum of all new native units issued that day.
IssTotUSD	The sum USD value of all new native units issued that day.
NVTAdj	The ratio of the network value divided by the adjusted transfer value.
NVTAdj90	The ratio of the network value to the 90-day moving average of the adjusted transfer value.

PriceBTC	The fixed closing price of the asset as of 00:00 UTC the following day denominated in BTC.
PriceUSD	The fixed closing price of the asset as of 00:00 UTC the following day denominated in USD.
ROI1yr	Return on investment 1 year
ROI30d	Return on investment 30 days
SplyCur	The sum of all native units ever created and currently visible on the ledger (i.e., issued) as of that day.
SplyExpFut10yrCMBI	This metric is an improved version of the legacy SplyExpFut10yr metric as it is better suited for use-cases that require a high degree of standardization, such as multi-asset indexes.
SplyFF	The sum of all native units ever created and visible on the ledger excluding native units held closely by company insiders, controlling investors and long term strategic holders as of that day
TxCnt	The sum count of transactions that day.
TxTfrCnt	The sum count of tranfers that interval(1 day in our case).
TxTfrValAdjNtv	The sum of native units transferred that day removing noise and certain artifacts.
TxTfrValAdjUSD	The USD value of the sum of native units transferred that day removing noise and certain artifacts.
TxTfrValMeanNtv	The sum value of native units transferred divided by the count of transfers between distinct addresses that interval.
TxTfrValMeanUSD	The sum USD value of native units transferred divided by the count of transfers ) that day.
TxTfrValMedNtv	The median count of native units transferred per transfer that day.
TxTfrValMedUSD	The median USD value transferred per transfer that day.
TxTfrValNtv	The sum of native units transferred that day.
TxTfrValUSD	The sum USD value of all native



	units transferred that day.
VtyDayRet180d	The 180 days volatility, measured as the deviation of log returns
VtyDayRet30d	The 30 days volatility, measured as the deviation of log returns
VtyDayRet60d	The 60 days volatility, measured as the deviation of log returns

Таблица 2: Используемые фундаментальные показатели

### 4.3. Макроэкономические показатели

Использование макроэкономических показателей широко применяется в области предсказания финансовых рядов, при этом они оказывают на них непосредственное влияние. В частности, это правило касается и рынка криптовалют [9]. Биткоин - неоднозначный актив, в нем есть как признаки биржевого товара, так и признаки обычной (фиатной) валюты [24]. Этот факт позволяет нам предположить, что в предсказании доходности криптовалют, нам могут помочь макроэкономические показатели, отвечающие за валютный рынок, а так как рассматриваются только пары криптовалюта/доллар, то ограничимся:

- DXY - индекс доллара. Показатель его силы или слабости
- LIBORUSD - лондонская межбанковская ставка предложения. Это средневзвешенная процентная ставка по кредитам на Лондонском межбанковском валютном рынке, рассчитана в долларах.

Также рассмотрим еще несколько важных показателей, которые показывают общее состояние рынка:

- SP500 - Индекс фондового рынка США состоящий из 500 компаний.
- VIX - Индекс волатильности S&P500
- GOLD - Цена на золото

## 5. Архитектуры нейронных сетей

### 5.1. Multi-layer perceptron

Многослойный перцептрон - классическая архитектура нейронной сети. В базовом исполнении состоит из: входного слоя нейронов(input layer), скрытого слоя нейронов(hidden layer) и выходного слоя нейронов(output layer). Причем связи между слоями направлены в одну сторону, так что эта нейронная сеть относится к классу нейронных сетей прямого пространства. Снизу на рисунке 1 как раз она представлена (картинка взята из [25]). Для обучения же этой сети можно использовать разнообразные алгоритмы, но самый распространенный - алгоритм обратного распространения ошибки, основанный на градиентном спуске [25]

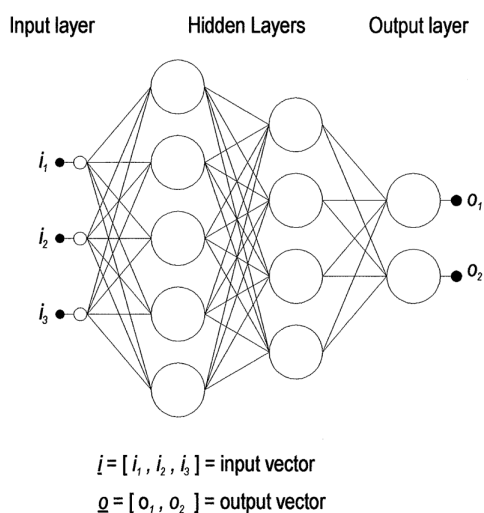


Рис. 1: MLP

### 5.2. LSTM[1]

Long short-term memory - нейронная сеть, которая относится к классу рекуррентных нейронных сетей, используется преимущественно в задачах, связанных с глубоким обучением, в том числе и с предсказанием финансовых временных рядов. Архитектура этой сети сильно сложнее, чем классического многослойного перцептрона. Далее рассмотрим ее подробнее. Все следующие формулы и изображения были взяты из [26]

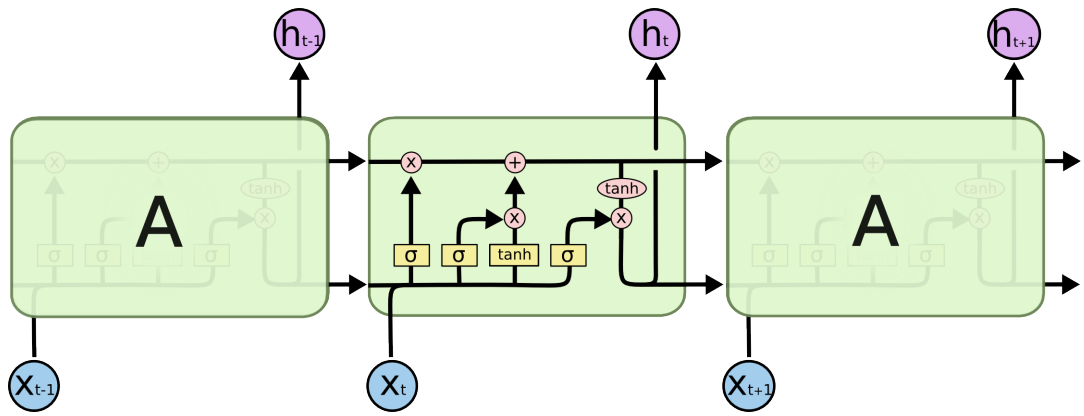


Рис. 2: LSTM-chain

Архитектура LSTM сети выглядит как цепь, состоящая из нескольких модулей, каждый из которых состоит из 4х слоев, которые взаимодействуют друг с другом определенным образом. Основное отличие этой сети от того же самого MLP в том, что у нас есть возможность передать состояние из одного модуля в следующий, при этом каждый модуль может удалить (забыть) часть этой информации, либо же наоборот что-то туда добавить, именно это и создает эффект ”памяти”

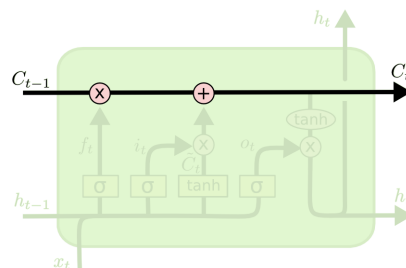


Рис. 3: Cell state C

Рассмотрим далее, как именно работает LSTM сеть, на примере полного пути данных через LSTM модуль.

В самом начале модуля мы попадаем в узел забвения (forget gate), который описывается следующей формулой:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

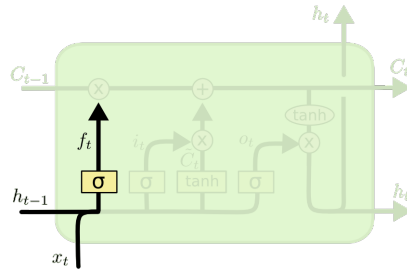


Рис. 4: LSTM-1st step

Следующий наш шаг будет состоять из двух. Сначала идет узел входного слоя (input gate layer).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

после же идет tanh слой

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

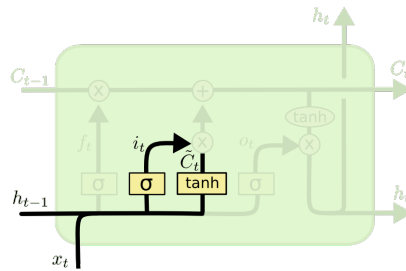


Рис. 5: LSTM-2d step

Следующим шагом мы обновляем старое состояние модуля на новое, для чего применим результаты первого и второго этапа к  $C_{t-1}$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

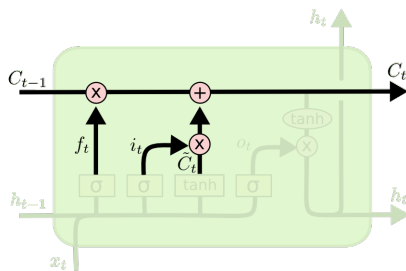


Рис. 6: LSTM-3d step

Последний этап - выходной узел (output gate). В нем мы получаем выходное внутреннее состояние ячейки.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

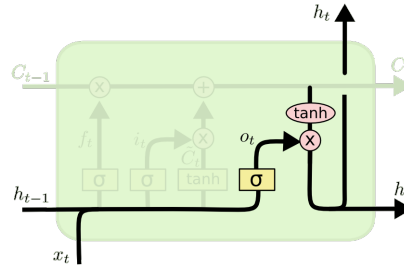


Рис. 7: LSTM-4th step

LSTM сеть отлично работает с данными, где надо запоминать какую-то информацию на долгий срок. Также она не так сильно подвержена проблеме исчезающего градиента, в отличие от обычной RNN [27]

### 5.3. GRU[2]

GRU(Gated Recurrent Unit) - еще одна сеть из семьи рекуррентных нейронных сетей. Фактически является пересмотром архитектуры LSTM в сторону ее упрощения. В ней входной узел и узел забвения объединены в один узел - узел обновления (update gate), также еще есть несколько изменений во внутренней архитектуре. GRU модуль можно описать следующими формулами:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

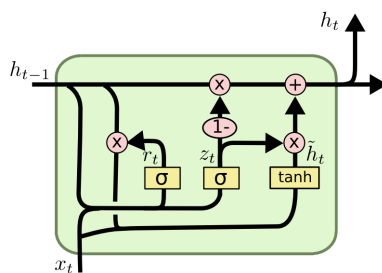


Рис. 8: GRU

К основным плюсам GRU по сравнению с LSTM относится более быстрое обучение, так как архитектура GRU проще, и, как следствие там меньше различных операций. При этом GRU сети могут работать лучше LSTM на тех же данных [28].

## 5.4. biLSTM[3]

Еще одна архитектура, которая используется в задаче анализа временных рядов, в том числе в предсказании цены на криптовалюты [7]. Двухнаправленный LSTM или biLSTM - модель, которая состоит из двух LSTM: один принимает входные данные в прямом направлении, а другой - в обратном. Таким образом получается, что объём информации для обучения увеличивается в 2 раза, что может положительно повлиять на итоговую ошибку алгоритма.

## 5.5. sLSTM

Stacked LSTM (многослойная LSTM сеть). Вид модели в которой используется не один LSTM слой, а сразу несколько. В результате получается "глубокая" нейронная сеть. Так как количество параметров сети вырастает, то есть вероятность, что итоговая модель будет точнее простой однослойной LSTM. Далее это будет проверено

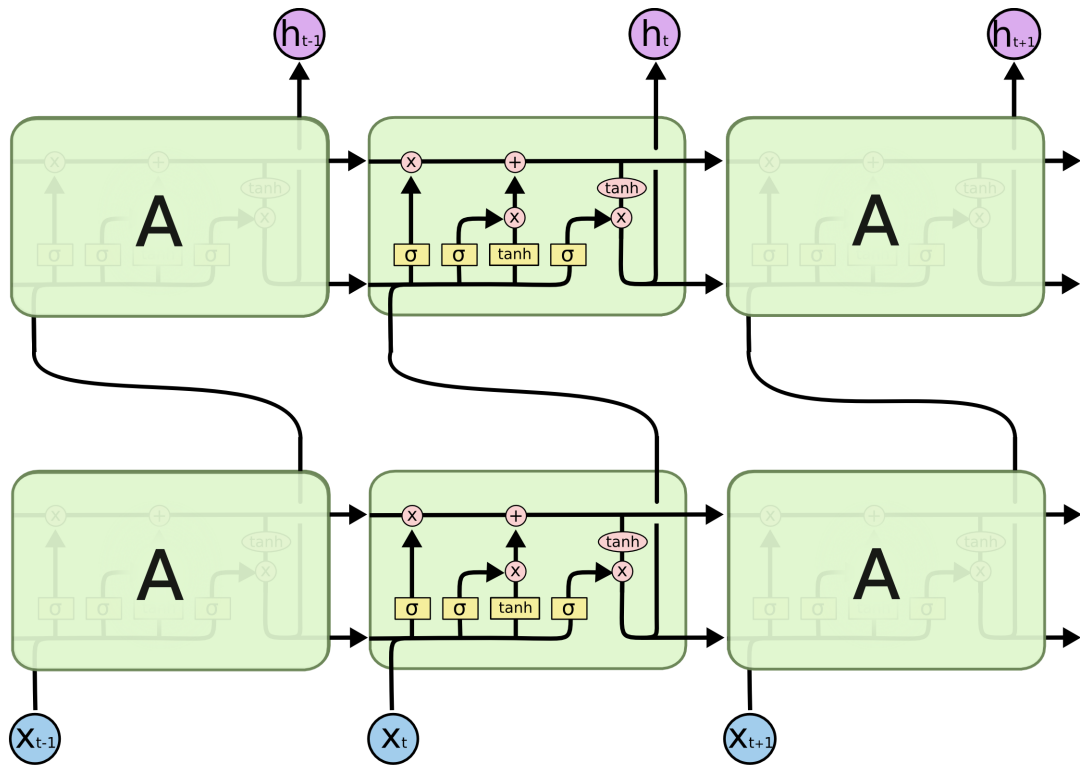


Рис. 9: LSTM-stacked

## 6. Методология

### 6.1. Подготовка данных

#### 6.1.1. Нормализация

Все признаки итогового набора данных лежат в различных интервалах, что мешает корректной работе модели. Чтобы решить эту проблему достаточно нормализовать все признаки. Для этого для каждого столбца итогового датасета достаточно применить формулу:

$$X_{std} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

где  $X$  - исходное значение (до нормализации),

$X_{std}$  - после

$\min(X)$ ,  $\max(X)$  - минимальное и максимальное значение признака соответственно.

#### 6.1.2. Уменьшение количества входных признаков

Существует два основных метода для сокращения количества входных признаков. Первый - создание по исходному набору признаков другого набора, с меньшей размерностью. Второй - выбор признаков из исходного набора, которые оказывают наибольшее влияние. К преимуществам второго метода над первым, прежде всего, относится то, что он сохраняет всю информацию о входных признаках и никак их не модернизирует (в первом методе работать приходится с "синтетическими" признаками). Согласно [29] выбор признаков ускоряет работу алгоритмов машинного обучения, при этом почти не понижает точности классификации (в статье решалась именно задача классификации).

Было решено для этой задачи решили использовать state of the art метод SHAP (SHapley Additive exPlanation) 2017г. [30], который хорошо показал себя в работе с финансовыми рядами [31]. В работе использовалась библиотека shap [32] для языка python. Обучим однослойную LSTN сеть (см LSTM[1]) по датасету технических и фундаментальных



признаков по отдельности, чтобы выбрать только те, которые оказывают наибольшее влияние на итоговый результат, макроэкономические признаки трогать не будем, так как их количество небольшое.

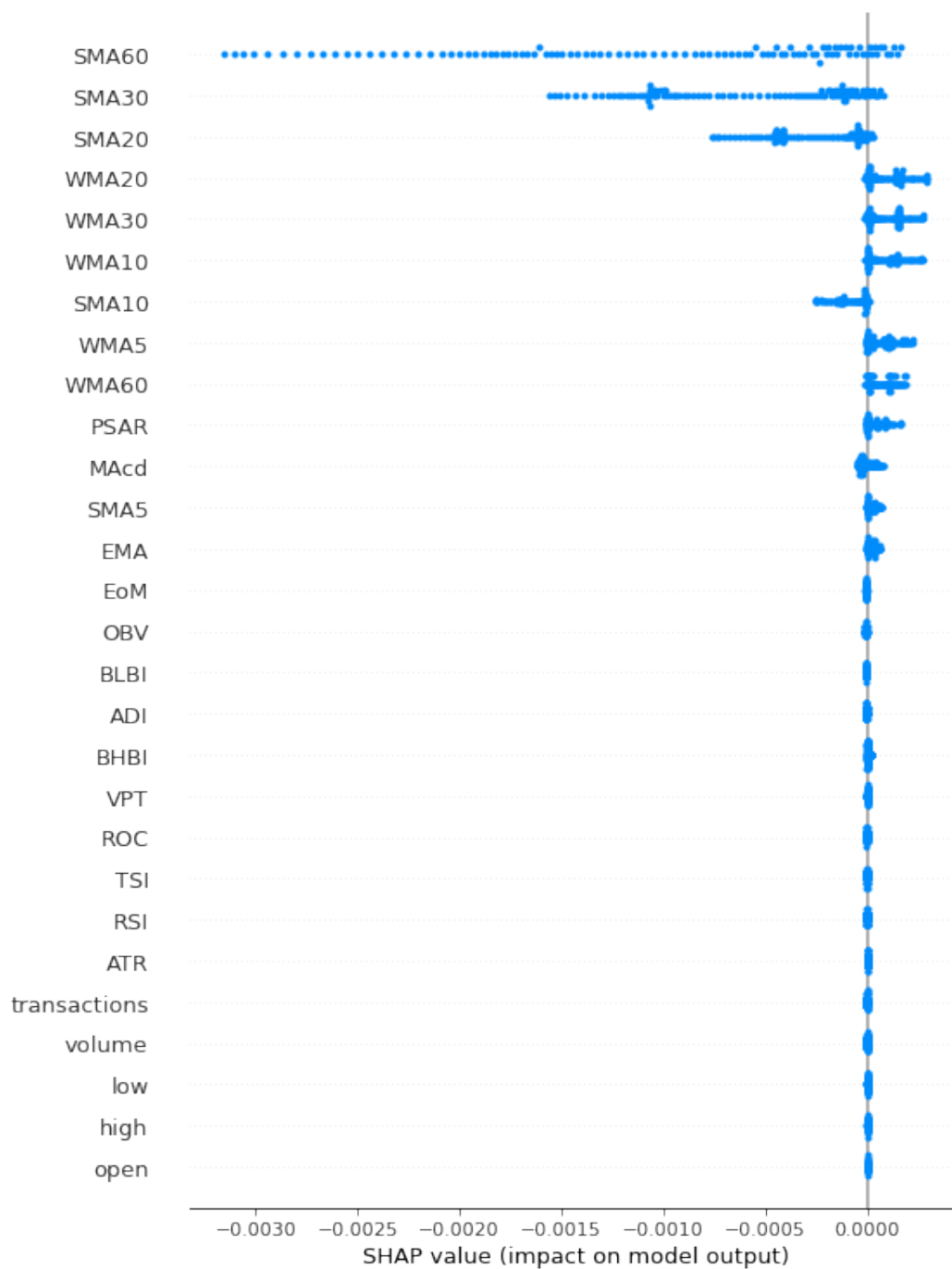


Рис. 10: Technical indicators impact

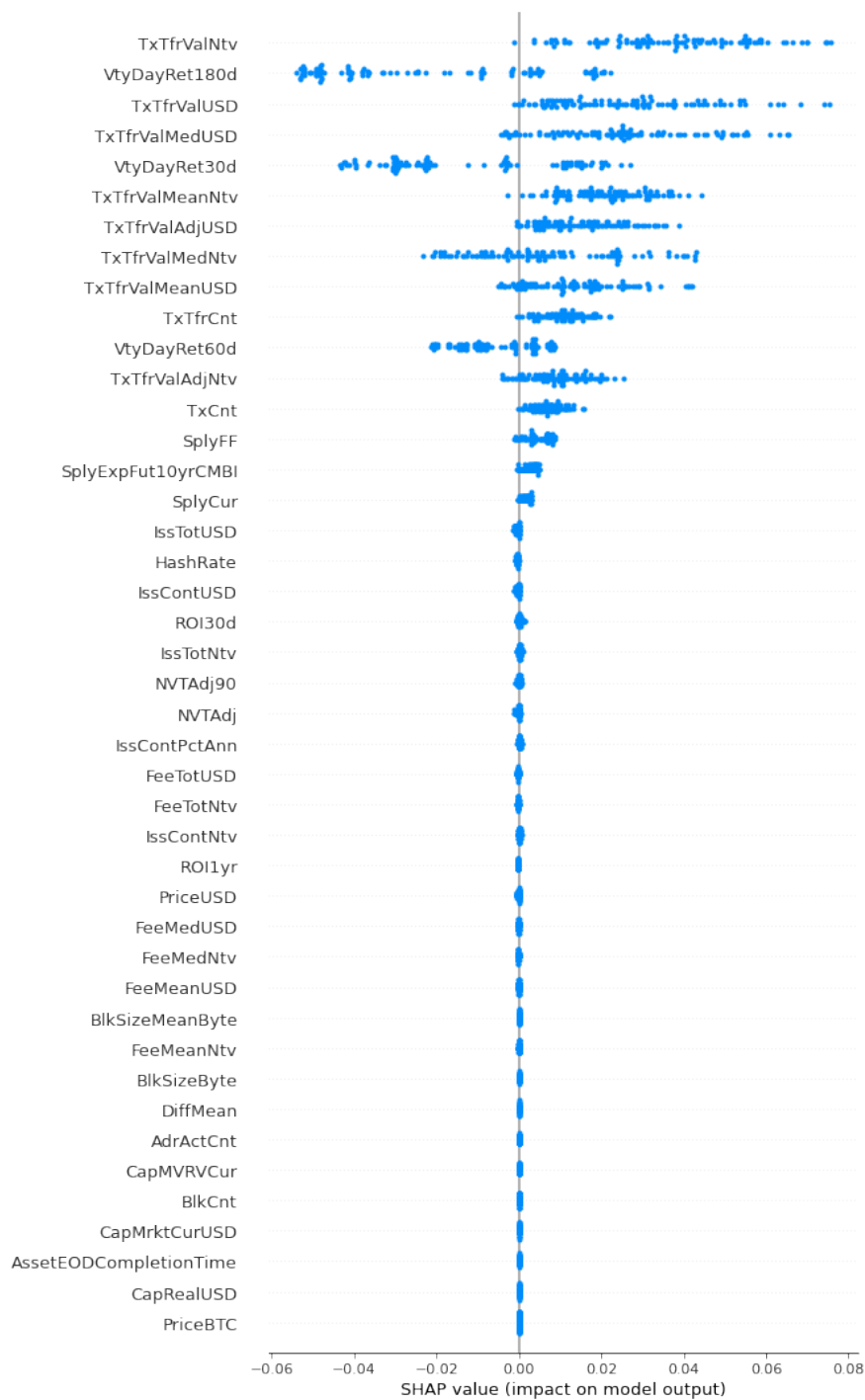


Рис. 11: Fundamental indicators impact

Как видно на Рис. 10, Рис. 11, далеко не все факторы имеют достаточное влияние на выходной результат, поэтому из технических индикаторов оставим только: все SMA, все WMA, PSAR, MACD, EMA. Из фундаментальных: 16 первых с Рис.11. В качестве проверки работоспособности этого метода выбора признаков, попробуем обучить наши модели для криптовалюты Dash и Bitcoin. В первом случае будем

использовать все технические индикаторы (23 признака), а во втором только те, что оказывают наибольшее влияние (13 признаков).

	23 Technical features	13 features selected by SHAP
LSTM	0.058	0.058
GRU	0.057	0.058
biLSTM	0.056	0.058
sLSTM	0.058	0.065

Таблица 3: RMSE for different technical datasets for Dash coin

	23 Technical features	13 features selected by SHAP
LSTM	0.052	0.054
GRU	0.051	0.054
biLSTM	0.053	0.053
sLSTM	0.050	0.057

Таблица 4: RMSE for different technical datasets for Bitcoin

Тоже самое сделаем и для фундаментальных признаков:

	43 Fundamental features	16 features selected by SHAP
LSTM	0.057	0.058
GRU	0.056	0.058
biLSTM	0.057	0.058
sLSTM	0.052	0.058

Таблица 5: RMSE for different fundamental datasets for Dash coin

	43 Fundamental features	16 features selected by SHAP
LSTM	0.045	0.046
GRU	0.044	0.046
biLSTM	0.046	0.046
sLSTM	0.042	0.046

Таблица 6: RMSE for different fundamental datasets for Bitcoin

Для остальных криптовалют построим только LSTM модель.

	23 Technical features	13 technical features selected by SHAP
btc	0.052	0.054
xrp	0.042	0.042
dash	0.058	0.058
xmr	0.051	0.052
ltc	0.018	0.019

Таблица 7: RMSE for LSTM model on different technical datasets

	43 Fundamental features	16 fundamental features selected by SHAP
btc	0.045	0.046
xrp	0.040	0.043
dash	0.057	0.058
xmr	0.053	0.054
ltc	0.018	0.019

Таблица 8: RMSE for LSTM model on different fundamental datasets

Вывод. Применение метода SHAP для выбора признаков в нашей задаче вполне оправдано. Так как удалось существенно понизить размерность датасетов, при этом точность итоговых моделей не уменьшилась вовсе, либо уменьшилась незначительно.

### 6.1.3. Кросс-валидация

Кросс-валидация (перекрестная проверка) - механизм, применяемый для более точной проверки результатов работы предсказательной модели. Нельзя исключать того, что разбивая данные на обучающее множество и проверочное множество, можно попасть в "удачный" интервал, то есть модель покажет очень низкую ошибку на нем, но основа любого эксперимента - его повторяемость, поэтому и используется кросс-валидация. Этот процесс заключается в том, что надо разбить датасет на несколько частей (в нашем случае было использовано 3 части [10]), и каждая часть состоит из обучающего и проверяющего блока соответственно. Всего существует несколько возможных компоновок этих блоков [33], будем использовать самую популярную версию. См Рис.12

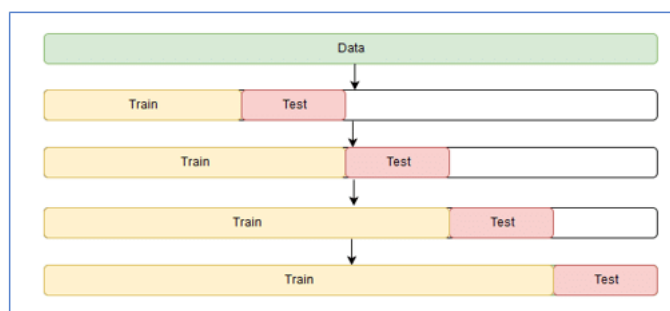


Рис. 12: Cross validation

В качестве итоговой ошибки модели возьмем среднюю ошибку на всех частях, как это было сделано в [10]. Размер учебной части датасета меняется в зависимости от этапа кросс-валидации ( $1/6$ ,  $3/6$ ,  $5/6$ ), размер проверяющей части постоянен и не меняется ( $1/6$ ,  $1/6$ ,  $1/6$ ).

## 6.2. Подготовка и запуск моделей

Каждая модель из главы 5 была реализована с помощью библиотеки Tensorflow [34] для языка программирования python3.9, на платформе Google Colab. Для обучения нейронных сетей необходимо выбрать оптимизационный алгоритм, в данный момент для таких задач обычно используют Adam [35]. Модели же будем тренировать с помощью

техники mini batch (вычисление ошибки происходит на каждом элементе датасета, но обновление весов модели производится только после вычисления всего пакета), которая является компромиссом между стандартным градиентным спуском и стохастическим градиентным спуском. Размер одного пакета установим на значение 64 [36]. Количество нейронов в однослойных моделях - 300 штук, количество нейронов в многослойных моделях 100 штук на слой. Тренировка будет проходить на протяжении 100 эпох. Для избегания переобучения и предотвращения коадпатации нейронов будем случайно выключать нейроны (dropout) с вероятностью 20

## 7. Результаты

См Таблица 9

См Рис. 13

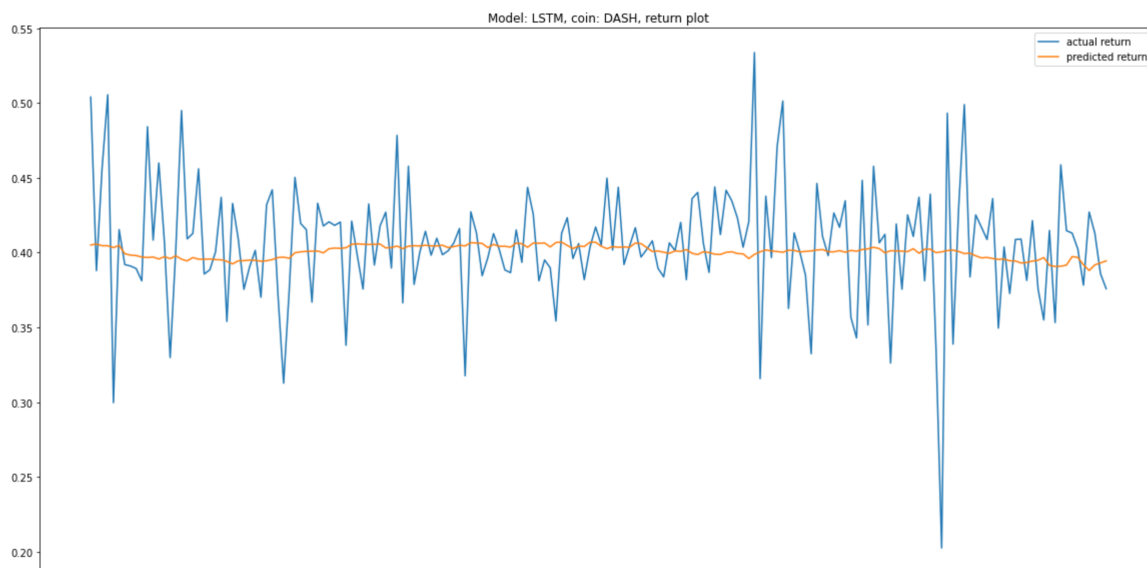


Рис. 13: Dash coin return example

### 7.1. Обсуждение результатов

- Различия итоговой ошибки у разных моделей - минимальные, это связано с природой данных, так как биржевые активы характеризуются высокой степенью эффективности и предсказать будущую доходность - очень сложная задача.
- Тем не менее все равно видна общая картина и понятно какие модели более эффективные, а какие менее. Аутсайдером стала MLP модель, что было ожидаемо, так как ее архитектура проще чем у рекуррентных сетей, также слабые результаты показала sLSTM модель, это скорее всего связано с недостатком данных для качественного обучения глубокой сети. Лучшие результаты показала GRU сеть.
- По датасетам лучшие результаты у датасета фундаментальных признаков. Объединение датасетов почти никак не повлияло на

model	coin	technical ind.	fundamental ind.	macroeconomics	united set Таб. 10
bilstm	btc	0.053	0.046	0.047	0.054
	dash	0.058	0.058	0.058	0.058
	ltc	0.020	0.019	0.020	0.020
	xmr	0.052	0.054	0.059	0.052
	xrp	0.042	0.043	0.046	0.043
gru	btc	0.054	0.046	0.047	0.053
	dash	0.058	0.058	0.058	0.058
	ltc	0.020	0.019	0.019	0.019
	xmr	0.052	0.054	0.055	0.052
	xrp	0.042	0.043	0.043	0.042
lstm	btc	0.054	0.046	0.047	0.053
	dash	0.058	0.058	0.058	0.058
	ltc	0.019	0.019	0.019	0.019
	xmr	0.052	0.054	0.059	0.051
	xrp	0.042	0.043	0.045	0.042
mlp	btc	0.057	0.055	0.053	0.053
	dash	0.067	0.061	0.065	0.060
	ltc	0.024	0.028	0.026	0.026
	xmr	0.059	0.055	0.059	0.052
	xrp	0.048	0.045	0.049	0.044
slstm	btc	0.057	0.046	0.049	0.053
	dash	0.065	0.058	0.063	0.058
	ltc	0.021	0.019	0.020	0.019
	xmr	0.059	0.054	0.060	0.051
	xrp	0.047	0.043	0.047	0.042
bilstm	mean	0.045	0.044	0.046	0.045
gru	mean	0.045	0.044	0.044	0.045
lstm	mean	0.045	0.044	0.046	0.045
mlp	mean	0.051	0.049	0.050	0.047
slstm	mean	0.050	0.044	0.048	0.045

Таблица 9: RMSE for all models and datasets

итоговую ошибку, что говорит о том, что нативное решение с объединением признаков почти не улучшает итоговую предсказательную способность.



## 8. Заключение

- В ходе работы были реализованы несколько основных моделей нейронных сетей для предсказания доходностей финансовых активов. Результаты их работы были сравнены друг с другом. На различных криптовалютах лучше работали различные архитектуры, но в среднем лучше всего себя показала GRU сеть. При этом MLP модель показала самые худшие результаты и проиграла всем остальным моделям
- Произведено сравнение моделей, обученных на разных датасетах: технических индикаторов, фундаментальных показателях и макроэкономических показателях. Также был собран итоговый датасет, который включал в себя все прочие датасеты.
- В ходе работы была показана эффективность использования алгоритма уменьшения размерности исходных данных SHAP [30], в результате работы которого были убраны технические признаки, которые использовались в работах [11][17] и просто замедляли процесс обучения нейронных сетей, не улучшая при этом результат. При этом после уменьшения кол-ва признаков в 2-3 раза итоговая ошибка нейронных сетей изменялась незначительно.

## Список литературы

- [1] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.
- [2] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” 2014.
- [3] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional lstm and other neural network architectures,” *Neural Networks*, vol. 18, no. 5, pp. 602–610, 2005. IJCNN 2005.
- [4] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” <http://bitcoin.org/bitcoin.pdf>,” 2008.
- [5] “Cryptocurrency Prices, Charts And Market Capitalizations,” <https://coinmarketcap.com/>.”
- [6] E. Karakoyun and A. Cibikdiken, “Comparison of arima time series model and lstm deep learning algorithm for bitcoin price forecasting,” in *The 13th Multidisciplinary Academic Conference in Prague*, vol. 2018, pp. 171–180, 2018.
- [7] T. Zoumpikas, E. Houstis, and M. Vavalis, “Eth analysis and predictions utilizing deep learning,” *Expert Systems with Applications*, vol. 162, p. 113866, 2020.
- [8] A. Dutta, S. Kumar, and M. Basu, “A gated recurrent unit approach to bitcoin price prediction,” *Journal of Risk and Financial Management*, vol. 13, p. 23, Feb 2020.
- [9] S. Corbet, C. Larkin, B. M. Lucey, A. Meegan, and L. Yarovaya, “The impact of macroeconomic news on bitcoin returns,” *The European Journal of Finance*, vol. 26, no. 14, pp. 1396–1416, 2020.
- [10] Z. Qiang, “Bitcoin High-Frequency Trend Prediction with Convolutional and Recurrent Neural Networks,” stanford winter report 70308950, Dstanford, Mar. 2021.

- [11] S. Alonso-Monsalve, A. L. Suárez-Cetrulo, A. Cervantes, and D. Quintana, “Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators,” *Expert Systems with Applications*, vol. 149, p. 113250, 2020.
- [12] W. Huang, K. K. Lai, Y. Nakamori, S. Wang, and L. Yu, “Neural networks in finance and economics forecasting,” *International Journal of Information Technology & Decision Making*, vol. 06, no. 01, pp. 113–140, 2007.
- [13] C. J. Neely, D. E. Rapach, J. Tu, and G. Zhou, “Forecasting the Equity Risk Premium: The Role of Technical Indicators,” *Management Science*, vol. 60, pp. 1772–1791, July 2014.
- [14] M. Hansson, “On stock return prediction with lstm networks,” 2017. Student Paper.
- [15] D. C. Yildirim, I. H. Toroslu, and U. Fiore, “Forecasting directional movement of forex data using lstm with technical and macroeconomic indicators,” *Financial Innovation*, vol. 7, p. 1, Jan 2021.
- [16] Y. Chen and Y. Hao, “Integrating principle component analysis and weighted support vector machine for stock trading signals prediction,” *Neurocomputing*, vol. 321, pp. 381–402, 2018.
- [17] W. Wang, W. Li, N. Zhang, and K. Liu, “Portfolio formation with preselection using deep learning from long-term financial data,” *Expert Systems with Applications*, vol. 143, p. 113042, 2020.
- [18] M. Nabipour, P. Nayyeri, H. Jabani, S. S., and A. Mosavi, “Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis,” *IEEE Access*, vol. 8, pp. 150199–150212, 2020.
- [19] Z. Hu, Y. Zhao, and M. Khushi, “A survey of forex and stock price prediction using deep learning,” *Applied System Innovation*, vol. 4, p. 9, Feb 2021.

- [20] “<https://technical-analysis-library-in-python.readthedocs.io/en/latest/>.”
- [21] T. Segal, “Fundamental analysis <https://www.investopedia.com/terms/f/fund> Apr 2021.
- [22] Y. Song, J. W. Lee, and J. Lee, “A study on novel filtering and relationship between input-features and target-vectors in a deep learning model for stock price prediction,” *Applied Intelligence*, vol. 49, pp. 897–911, Mar 2019.
- [23] “<https://coinmetrics.io/>,” Feb 2021.
- [24] D. Baur, K. Hong, and A. Lee, “Bitcoin: Medium of exchange or speculative assets?,” *Journal of International Financial Markets, Institutions and Money*, vol. 54, 12 2017.
- [25] M. Gardner and S. Dorling, “Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences,” *Atmospheric Environment*, vol. 32, no. 14, pp. 2627–2636, 1998.
- [26] C. Olah, “Understanding lstm networks,” 2015.
- [27] S. Hochreiter, “The vanishing gradient problem during learning recurrent neural nets and problem solutions,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, pp. 107–116, 04 1998.
- [28] P. T. Yamak, L. Yujian, and P. K. Gadosey, “A comparison between arima, lstm, and gru for time series forecasting,” in *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, ACAI 2019, (New York, NY, USA), p. 49–55, Association for Computing Machinery, 2019.
- [29] I. Rodriguez-Lujan, R. Huerta, C. Elkan, and C. S. Cruz, “Quadratic programming feature selection,” *J. Mach. Learn. Res.*, vol. 11, p. 1491–1516, Aug. 2010.

- [30] S. Lundberg and S. Lee, “A unified approach to interpreting model predictions,” *CoRR*, vol. abs/1705.07874, 2017.
- [31] X. Xiaomao, Z. Xudong, and W. Yuanfang, “A comparison of feature selection methodology for solving classification problems in finance,” *Journal of Physics: Conference Series*, vol. 1284, p. 012026, aug 2019.
- [32] “Shap python library, <https://github.com/slundberg/shap>.”
- [33] V. Cerqueira, L. Torgo, and I. Mozetič, “Evaluating time series forecasting models: an empirical study on performance estimation methods,” *Machine Learning*, vol. 109, pp. 1997–2028, Nov 2020.
- [34] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. Software available from tensorflow.org.
- [35] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2017.
- [36] K. Chen, Y. Zhou, and F. Dai, “A lstm-based method for stock returns prediction: A case study of china stock market,” in *2015 IEEE International Conference on Big Data (Big Data)*, pp. 2823–2824, 2015.

## Приложение

dxy	liborUSD	sp500
vix	gold	open
high	low	close
volume	transactions	SMA60
SMA30	SMA20	SMA10
SMA5	WMA60	WMA30
WMA20	WMA10	WMA5
PSAR	MAcd	EMA
SplyFF	TxTfrValNtv	VtyDayRet180d
TxTfrValUSD	TxTfrValMedUSD	VtyDayRet30d
TxTfrValMeanNtv	TxTfrValAdjUSD	TxTfrValMedNtv
TxTfrValMeanUSD	TxTfrCnt	VtyDayRet60d
TxTfrValAdjNtv	TxCnt	SplyExpFut10yrCMBI
SplyCur		

Таблица 10: important features dataset