

# ПРЕДВИЋАЊЕ ТРЕНДА КРИПТОВАЛУТА НА ОСНОВУ ОБЈАВА НА ДРУШТВЕНИМ МРЕЖАМА

*Увод у неуронске мреже и Моделовање  
финансијских тржишта*

Јанко Гашић 134/2016

Иван Прелић 421/2016

Никола Ђорђиески 4/2017

Анђелија Милутиновић 12/2017


Никола Керезовић 1053/2017

# Предмет



# Циль





# Прикупљање и претпроцесирање података

# Прикупљање података са друштвене мреже Редит

Период: 1. децембар 2020 – 1. мај 2021.

Криптовалуте: Bitcoin, Ethereum, Ripple

Програмски језик: Python

АПИ: Pushshift, PRAW



# Структура података

100 најпопуларнијих објава у дану  
Одвојене табеле за сваку од криптовалута  
Свака од табела има 13 колона

Биткоин - 13099 опсервација  
Етериум - 11336 опсервација  
Рипл - 4499 опсервација



# Структура података - атрибути

Post id - јединствена идентификација поста

Title - наслов објаве

Date and Time - датум и време када је објава постављена на Редит

Date - датум када је објава постављена на Редит

Time - време када је објава постављена на Редит

Hour - сат када је објава постављена на Редит

Upvotes - укупан број позитивних гласова на објави

Compound Sentiment - укупан сентимент добијен помоћу VADER лексикона

Voted Score - коначни сентимент

Positive Sentiment - укупан број позитивних речи у једној објави

Negative Sentiment - укупан број негативних речи у једној објави

Sentiment Score - разлика између позитивних и негативних речи у једној објави

# Начин прикупљања података



# Неопходне библиотеке

```
import requests
import json
import datetime
import time
import re
import pandas
import openpyxl
import praw
import nltk

nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer
```

Requests

JSON

DateTime

Time

Re

Pandas

OpenPyxl

Praw

Nltk

SentimentIntensityAnalyzer

# Главна функција

```
def subreddit_pushshift_time_top_retriever(subreddit_name,  
start_time, end_time, last_month_mode=False)
```

Пример параметара:

- `subreddit_name = "Bitcoin"`
- `start_time = 1606777200`
- `end_time = 1606863599`
- `last_month_mode = False`

# Главна функција - упит

Упит помоћу којег АПИ прикупља податке:

*api\_search\_query*

# Главна функција - прикупљање података

```
if len(list_of_ids) != 0: # If we don't have ANY data, just return an empty dictionary.
    # Get info for each object. 'info()' is quite fast and accepts a list of fullname IDs.
    reddit_submissions = reddit.info(fullnames=list_of_ids)
    for submission in reddit_submissions:
        # post_score = submission.score
        final_dictionary[post_id] = {}
        final_dictionary[post_id]['id'] = submission.id
        final_dictionary[post_id]['score'] = submission.score
        final_dictionary[post_id]['title'] = submission.title
        final_dictionary[post_id]['link'] = submission.permalink
        final_dictionary[post_id]['created_utc'] = int(submission.created_utc)
        final_dictionary[post_id]['author'] = submission.author
        final_dictionary[post_id]['num_comments'] = submission.num_comments
        final_dictionary[post_id]['self_text'] = submission.selftext
        final_dictionary[post_id]['compound_sentiment'] = 0
        final_dictionary[post_id]['voted_score'] = 0
        final_dictionary[post_id]['positive_sentiment'] = 0
        final_dictionary[post_id]['negative_sentiment'] = 0
        final_dictionary[post_id]['sentiment_score'] = final_dictionary[post_id]['positive_sentiment'] - \
            final_dictionary[post_id]['negative_sentiment']
        post_id += 1
```

# Главна функција - сентимент анализа

```
j = 0
while j < 100:

    # TITLE ANALYSIS
    title = final_dictionary[j]['title']
    title_score = sid.polarity_scores(title)['compound']
    # print(title_score)

    text = final_dictionary[j]['self_text']
    text_score = sid.polarity_scores(text)['compound']
    # print(text_score)

    text_and_title_score = title_score + text_score
    # print(text_and_title_score)

    post_score = final_dictionary[j]['score']
    # print(post_score)
    voted_score = text_and_title_score * (1 + post_score * 0.2)
    # print(voted_score)

    final_dictionary[j]['compound_sentiment'] = text_and_title_score
    final_dictionary[j]['voted_score'] = voted_score
```

# Главна функција - сентимент анализа

```
title2 = re.sub('[.,!?:;]', ' ', final_dictionary[j]['title'])
words_title = re.findall(r"[\w']+", title2)
for word in words_title:
    for positive_word in x:
        if positive_word == word.lower():
            positive_counter = positive_counter + 1
            break
    for negative_word in y:
        if negative_word == word.lower():
            negative_counter = negative_counter + 1
            break

# IF TEXT AVAILABLE TEXT ANALYSIS
if final_dictionary[j]['self_text'] != "":
    text = re.sub('[.,!?:;]', ' ', final_dictionary[j]['self_text'])
    words_text = re.findall(r"[\w']+", text)
    for word in words_text:
        for positive_word in x:
            if positive_word == word.lower():
                positive_counter = positive_counter + 1
                break
        for negative_word in y:
            if negative_word == word.lower():
                negative_counter = negative_counter + 1
                break
```

# Главна функција - упис у речник

```
n = 0
while n < 100:
    date_date = time.strftime('%Y-%m-%d %H:%M:%S', time.localtime(final_dictionary[n]['created_utc']))

    ids.append(final_dictionary[n]['id'])
    scores.append(final_dictionary[n]['score'])
    titles.append(final_dictionary[n]['title'])
    authors.append(final_dictionary[n]['author'])
    positive_sentiments.append(final_dictionary[n]['positive_sentiment'])
    negative_sentiments.append(final_dictionary[n]['negative_sentiment'])
    sentiment_scores.append(final_dictionary[n]['sentiment_score'])
    comments.append(final_dictionary[n]['num_comments'])
    dates.append(date_date)
    compound_sentiments.append(final_dictionary[n]['compound_sentiment'])
    voted_scores.append(final_dictionary[n]['voted_score'])
    epoch_times.append(final_dictionary[n]['created_utc'])
    n += 1

return final_dictionary
```

# Позив главне функције

```
start = 1606777200 # 1.dec2020
end = 1606863599
p = 0

while p < 151:
    d = time.strftime('%Y-%m-%d %H:%M:%S', time.localtime(start))
    print(d)
    try:
        subreddit_pushshift_time_top_retriever('Bitcoin', start, end, last_month_mode=False)
    except:
        print("Greska 1")
    start += 86400
    end += 86400
    p += 1
```



# Позив главне функције

```
start = 1606777200 # 1.dec2020
end = 1606863599
p = 0

while p < 151:
    d = time.strftime('%Y-%m-%d %H:%M:%S', time.localtime(start))
    print(d)
    try:
        subreddit_pushshift_time_top_retriever('Bitcoin', start, end, last_month_mode=False)
    except:
        print("Greska 1")
    start += 86400
    end += 86400
    p += 1
```

# Чување добијених података

```
new_dataframe = pandas.DataFrame(  
    {  
        "Post id": ids,  
        "Epoch time": epoch_times,  
        "Title": titles,  
        "Date": dates,  
        "Upvotes": scores,  
        "Compound Sentiment": compound_sentiments,  
        "Voted Score": voted_scores,  
        "Positive Sentiment": positive_sentiments,  
        "Negative Sentiment": negative_sentiments,  
        "Sentiment Score": sentiment_scores,  
        "Username": authors,  
        "Num of comments": comments  
    }  
)  
  
writer = pandas.ExcelWriter('outputBITCOIN.xlsx')  
try:  
    new_dataframe.to_excel(writer, 'Sheet1')  
except:  
    print("Greska 2")  
writer.save()
```

# Анализа тржишног сентимента

# Рачунање сентимента

- Две улазне варијабле: Број гласова и сентимент текста
- Рачунање сентимента – *VADER* лексикон
- Излаз из методе за рачунање: позитивни, негативни, неутрални, сложени сентимент
- Формула:  $Voted\ score = Compound\ score \times (1 + Score \times 0.2)$ 
  - Score: Разлика између позитивних и негативних гласова
  - Compound Sentiment: Сложени сентимент
  - Voted Score: Коначни сентимент

# Спајање и груписање података

- Спајање података са Редита и цена криптовалута у један скуп
- Пивот табела за груписање података по сатима
- Нове колоне:

Row Labels	Count of Posts	Sum of Voted Score	BTC open value	BTC close value
1.12.2020	1	1.25554	19695.87	19565.47
1.12.2020 01:00	2	0.72744	19565.47	19605.75
1.12.2020 02:00	5	-0.10266	19605.75	19680.95
1.12.2020 03:00	5	1.40806	19680.96	19419.74
1.12.2020 04:00	4	2.72938	19419.73	19354.31
1.12.2020 05:00	7	1.87596	19352.64	19483.73
1.12.2020 06:00	5	2.53004	19483.73	19338.34

- Count of Posts
- Sum of Voted Score
- BTC open value
- BTC close value

# Претходни тренд и будућа промена

Нове колоне:

- Промена
- Знак промене
- Претходни тренд
- Знак претходног тренда
- Будућа промена
- Знак будуће промене

Рачунање претходног тренда:

$$T = \frac{cv_{end} - cv_{start}}{cv_{start}} \times 100\%$$

Рачунање будуће промене:

$$P = \frac{cv_{end2} - cv_{start2}}{cv_{start2}} \times 100\%$$

# Коначна табела

	Row Labels	Count of Posts	Sum of Voted Score	BTC open value	BTC close value	Promena	Znak	Prethodni Trend	Znak Prethodni	Buduca Promena	Znak Buduca
0	2020-12-01 00:00:00	1	1.25554	19695.87	19565.47	-130.4	-1	0	0	-0.00718	-1
1	2020-12-01 01:00:00	2	0.72744	19565.47	19605.75	40.28	1	0	0	-0.05378	-1
2	2020-12-01 02:00:00	5	-0.10266	19605.75	19680.95	75.2	1	0	0	-0.04681	-1
3	2020-12-01 03:00:00	5	1.40806	19680.96	19419.74	-261.22	-1	0	0	-0.00692	-1
4	2020-12-01 04:00:00	4	2.72938	19419.73	19354.31	-65.42	-1	0	0	-0.0047	-1
5	2020-12-01 05:00:00	7	1.87596	19352.64	19483.73	131.09	1	0	0	-0.02181	-1
6	2020-12-01 06:00:00	5	2.53004	19483.73	19338.34	-145.39	-1	0	0	-0.01448	-1
7	2020-12-01 07:00:00	3	14.39062	19338.33	19515.63	177.3	1	0	0	-0.0398	-1
8	2020-12-01 08:00:00	2	0.5058	19515.62	19466.99	-48.63	-1	0	0	-0.0204	-1
9	2020-12-01 09:00:00	1	0.27156	19467	19565	98	1	0	0	-0.02764	-1
10	2020-12-01 10:00:00	6	9.5668	19564.99	19739.51	174.52	1	0	0	-0.03552	-1

# Статистика

Улазне варијабле:

- Сентимент
- Претходни тренд

Излазна варијабла:

- **Знак будуће промене**



## Приказ тачно предвиђене позитивне будуће промене

	Sum of Voted Score	Previous Trend	Future Change
0	40.66146	0.24	0.65
1	1.73364	1.32	1.22
2	1.53638	1.84	1.32

## Приказ нетачно предвиђене неутралне будуће промене

	Sum of Voted Score	Previous Trend	Future Change
0	0	0	-2.73
1	0	0	-2.66
2	0	0	0.29

## Приказ тачно предвиђене будуће промене на основу знака коначног сентимента

	Sum of Voted Score	Previous Trend	Future Change
0	-0.10266	0	-4.68
1	1.70616	0	1.15
2	-1.1357	0	-2.24
3	0.69146	0	0.53
4	0.65328	0	0.82
5	2.86696	-3.92	1.54

## Приказ тачно предвиђене будуће промене на основу знака претходног тренда

	Sum of Voted Score	Previous Trend	Future Change
0	-0.30008	2.89	1.38
1	-1.2839	0.07	0.85
2	1.6838	-0.57	-0.91
3	0.10224	-1.09	-0.89
4	1.65134	-0.84	-2.11
5	0	-2.25	-0.81

# Метрике за све 3 криптовалите

	<i>TP</i>	<i>FP</i>	<i>TN</i>	<i>FN</i>	<i>TNe</i>	<i>FNe</i>	<i>TScore</i>	<i>TTrend</i>	<i>Ostali slučajevi</i>
<b>BTC</b>	0.127	0.121	0.016	0.029	0.000	0.034	0.326	0.124	0.224
<b>ETH</b>	0.120	0.128	0.014	0.026	0.000	0.063	0.316	0.146	0.187
<b>XRP</b>	0.076	0.069	0.032	0.035	0.000	0.125	0.249	0.161	0.252

# Дескриптивна статистика

<i>True Positive</i>		
<i>Average Of Sum Of Voted Score</i>	<i>Average Of Previous Trend</i>	<i>Average Of Future Change</i>
<b>16.450945</b>	<b>5.114215116</b>	<b>2.737761628</b>
<i>StdDev Of Sum Of Voted Score</i>	<i>StdDev Of Previous Trend</i>	<i>StdDev Of Future Change</i>
<b>49.0286512</b>	<b>3.948553925</b>	<b>2.562662369</b>
<i>False Positive</i>		
<i>Average Of Sum Of Voted Score</i>	<i>Average Of Previous Trend</i>	<i>Average Of Future Change</i>
<b>22.71852031</b>	<b>5.340552147</b>	<b>-1.952576687</b>
<i>StdDev Of Sum Of Voted Score</i>	<i>StdDev Of Previous Trend</i>	<i>StdDev Of Future Change</i>
<b>74.11606775</b>	<b>5.087514823</b>	<b>1.966596619</b>

# Дескриптивна статистика

<i>True Negative</i>		
<i>Average Of Sum Of Voted Score</i>	<i>Average Of Previous Trend</i>	<i>Average Of Future Change</i>
<b>-25.29898381</b>	<b>-4.407142857</b>	<b>-3.014285714</b>
<i>StdDev Of Sum Of Voted Score</i>	<i>StdDev Of Previous Trend</i>	<i>StdDev Of Future Change</i>
<b>97.221558</b>	<b>4.017998704</b>	<b>2.908245756</b>
<i>False Negative</i>		
<i>Average Of Sum Of Voted Score</i>	<i>Average Of Previous Trend</i>	<i>Average Of Future Change</i>
<b>-12.50138727</b>	<b>-5.706493506</b>	<b>2.586493506</b>
<i>StdDev Of Sum Of Voted Score</i>	<i>StdDev Of Previous Trend</i>	<i>StdDev Of Future Change</i>
<b>44.76848745</b>	<b>4.959226321</b>	<b>2.455869321</b>


# Дескриптивна статистика

<i>False Neutral</i>		
<i>Average Of Sum Of Voted Score</i>	<i>Average Of Previous Trend</i>	<i>Average Of Future Change</i>
<b>0</b>	<b>0</b>	<b>1.138817204</b>
<i>StdDev Of Sum Of Voted Score</i>	<i>StdDev Of Previous Trend</i>	<i>StdDev Of Future Change</i>
<b>0</b>	<b>0</b>	<b>4.321497287</b>

# Дескриптивна статистика

<i>True Prediction According To The Voted Score</i>		
<i>Average Of Sum Of Voted Score</i>	<i>Average Of Previous Trend</i>	<i>Average Of Future Change</i>
<b>16.20174865</b>	<b>-0.684084187</b>	<b>1.58817975</b>
<i>StdDev Of Sum Of Voted Score</i>	<i>StdDev Of Previous Trend</i>	<i>StdDev Of Future Change</i>
<b>105.542943</b>	<b>3.585954351</b>	<b>3.588164245</b>
<i>True Prediction According To The Previous Trend</i>		
<i>Average Of Sum Of Voted Score</i>	<i>Average Of Previous Trend</i>	<i>Average Of Future Change</i>
<b>7.308021138</b>	<b>0.281766467</b>	<b>-0.173023952</b>
<i>StdDev Of Sum Of Voted Score</i>	<i>StdDev Of Previous Trend</i>	<i>StdDev Of Future Change</i>
<b>75.83389324</b>	<b>5.344253757</b>	<b>4.232244893</b>





# Коришћене методе вештачке интелигенције

# Методе класификације

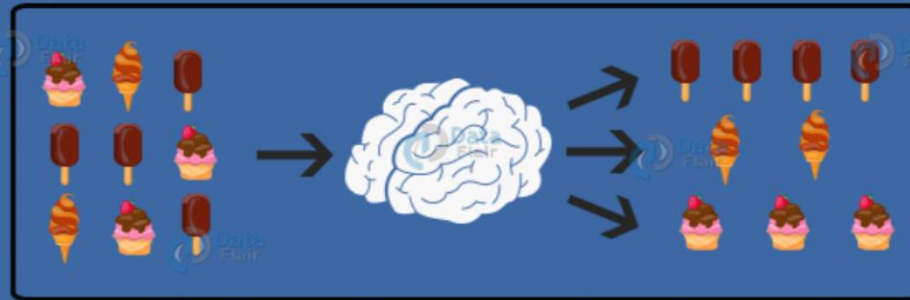
## Machine Learning Classification Algorithms



Logistic Regression

Naive Bayes

Decision Tree



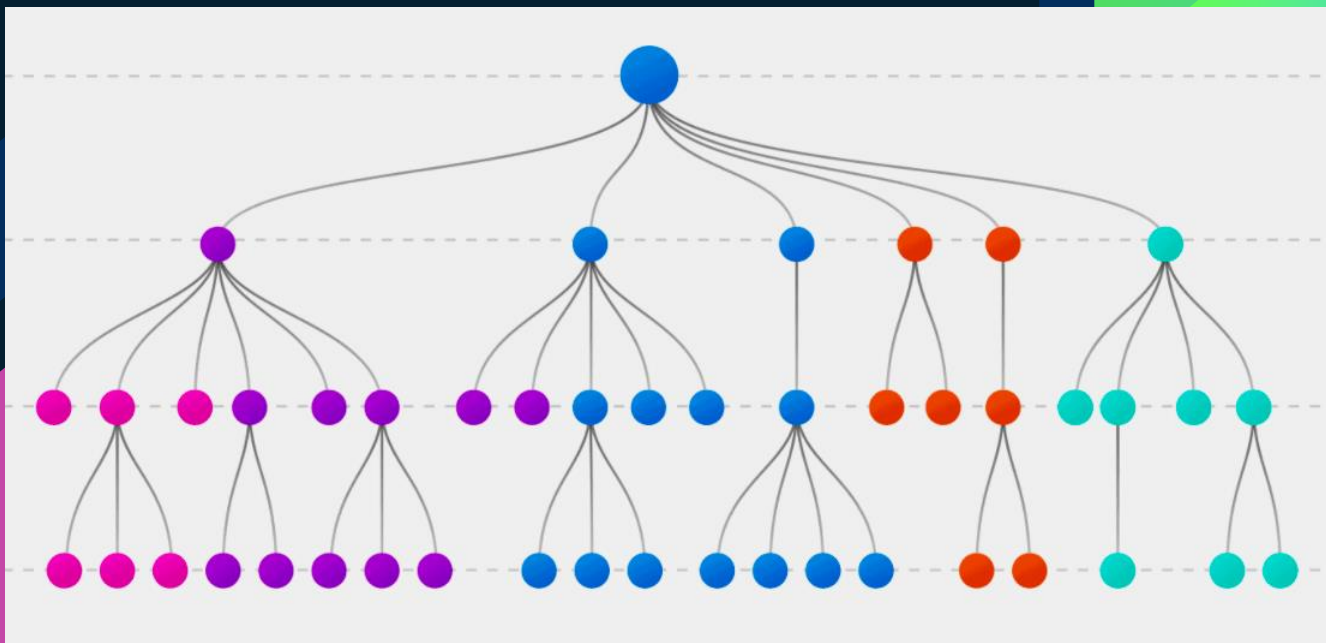
Support Vector Machines

Random Forest

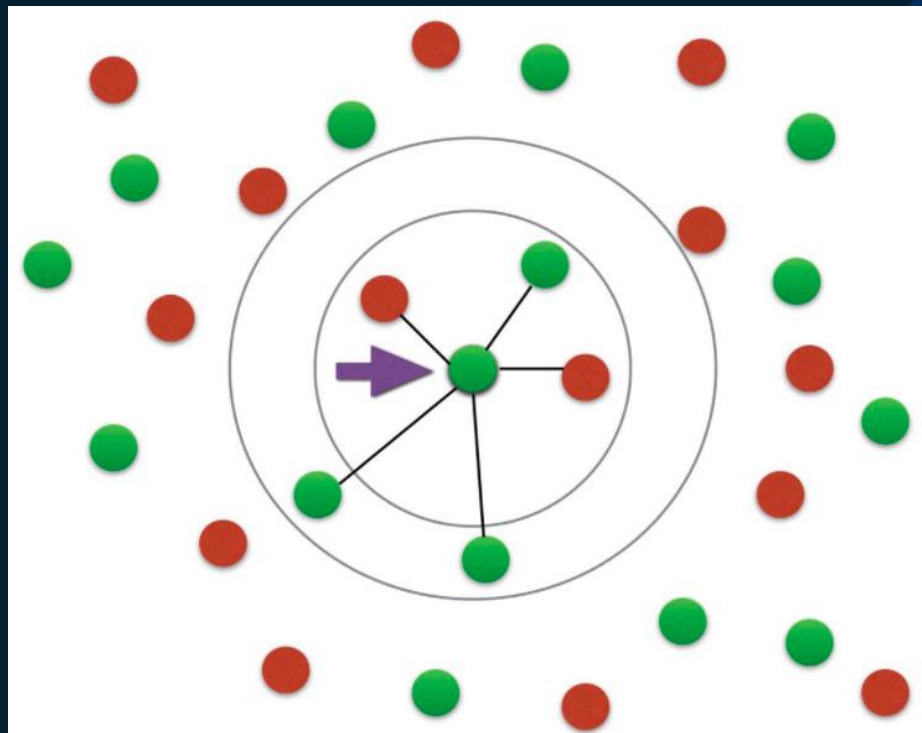
K-Nearest Neighbours



# Стабло одлучивања



# К-најближих суседа



# Резултати метода класификације

	BTC	XRP	ETH
Број погођених вредности	261	185	293
Укупан број вредности за предвиђање	532	396	549
Тачност	0.49	0.48	0.53

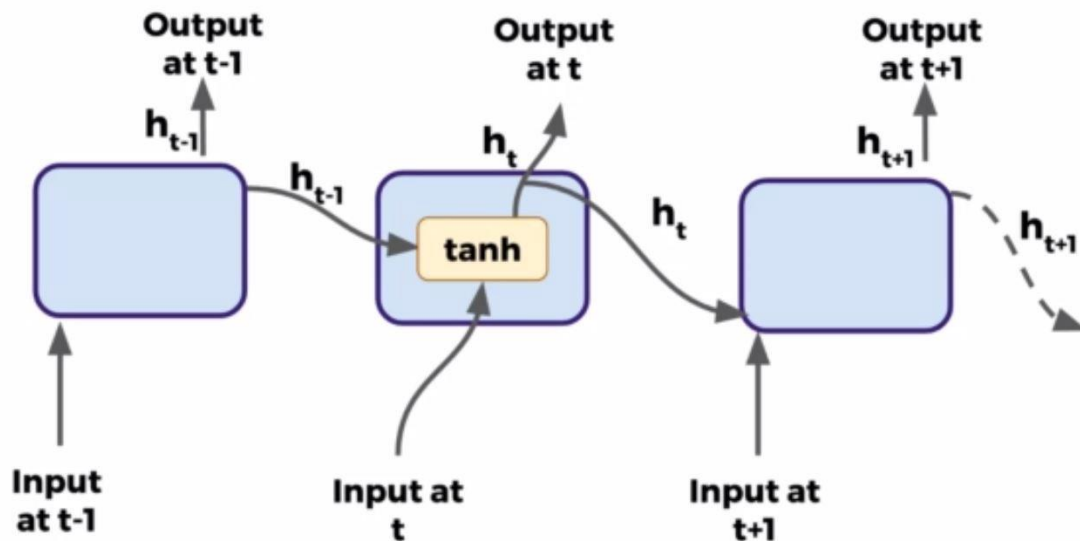
	BTC	XRP	ETH
Број погођених вредности	262	181	295
Укупан број вредности за предвиђање	532	396	549
Тачност	0.49	0.46	0.54

Стабло  
одлучивања

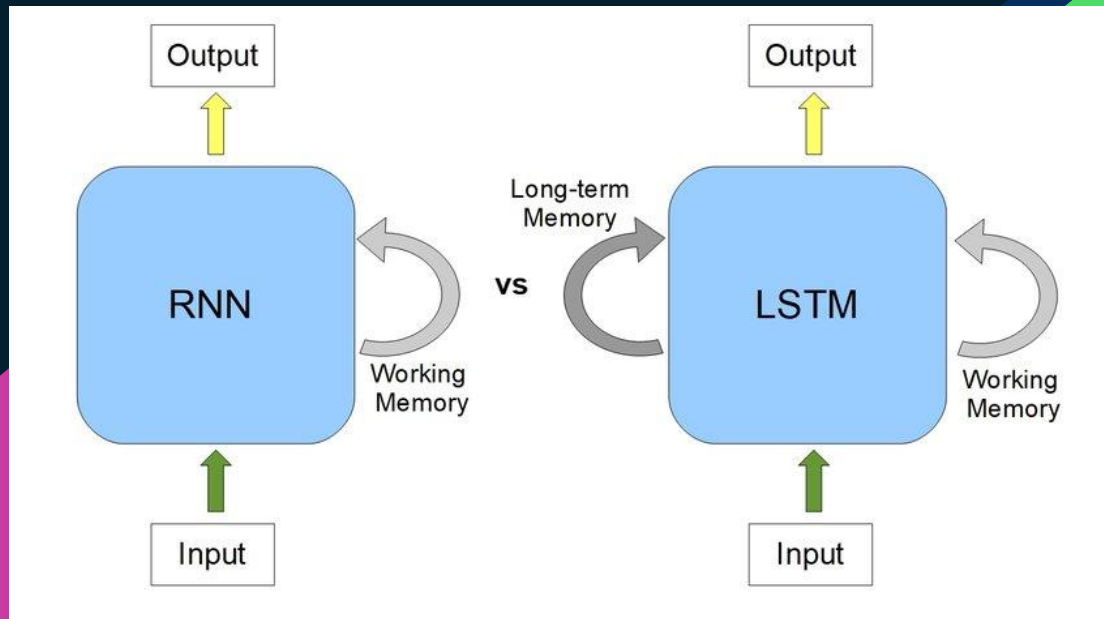
К-најближих  
суседа

# Рекурентне неуронске мреже

A typical RNN cell



# РНН и ЛСТМ (*Long short term memory*)



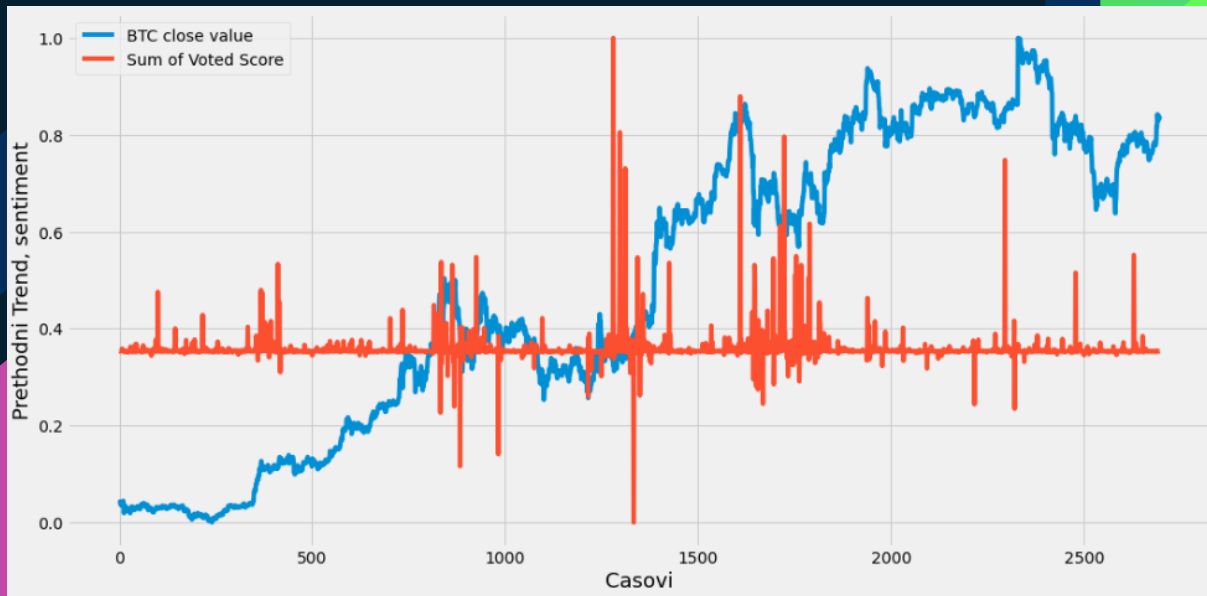
Можућност дугорочног памћења

# Примена ЛСТМ мреже

- Помоћу вредности сентимента и цене у претходна 24. часа предвиђамо вредност у наредном часу
- Тренинг 86% скупа података, тест 12%
- Два ЛСТМ и два Dense слоја са различитим бројем неурона
- Евалуација модела: корен средњег квадратног одступања



# Корелација цене Биткоина и сентимента

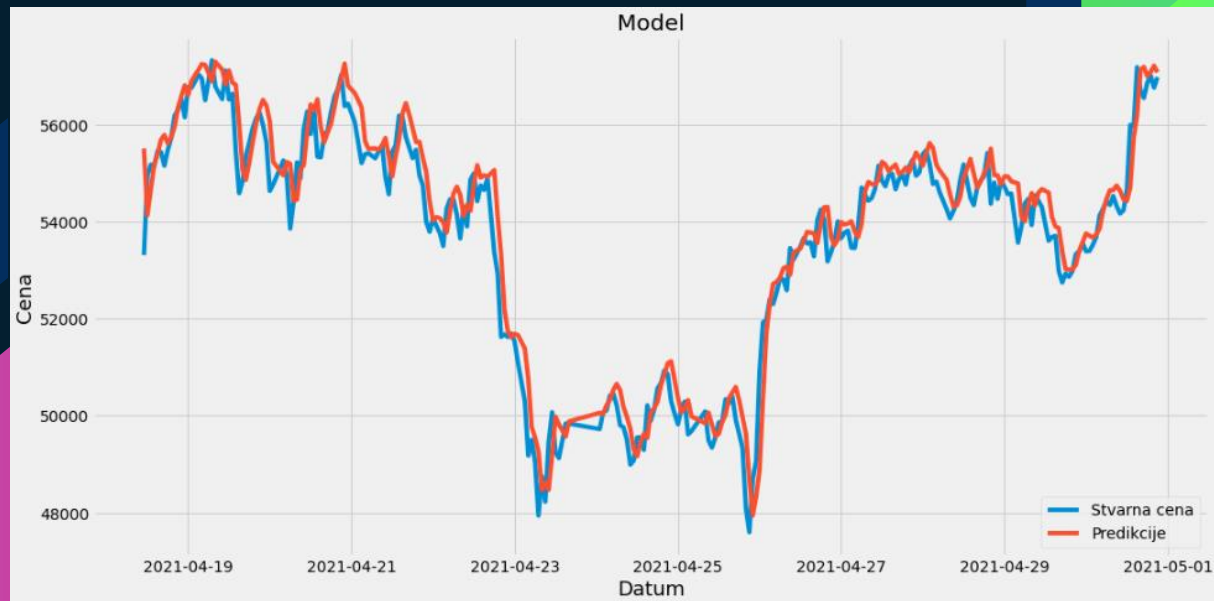


# Резултати

Epochs		1	10	100
Batch size				
1		270.67	439.97	768.75
10		117.67	40.62	288.21
100		73.62	108.86	149.55

Број слојева и неурона	Корен средње квадратне грешке
LSTM(10), LSTM(10), Dense(10), Dense(1)	628.88
LSTM(25), LSTM(25), Dense(10), Dense(1)	569.46
LSTM(50), LSTM(50), Dense(25), Dense(1)	40.62
LSTM(100), LSTM(100), Dense(50), Dense(1)	108.90
LSTM(200), LSTM(150), Dense(100), Dense(1)	181.18

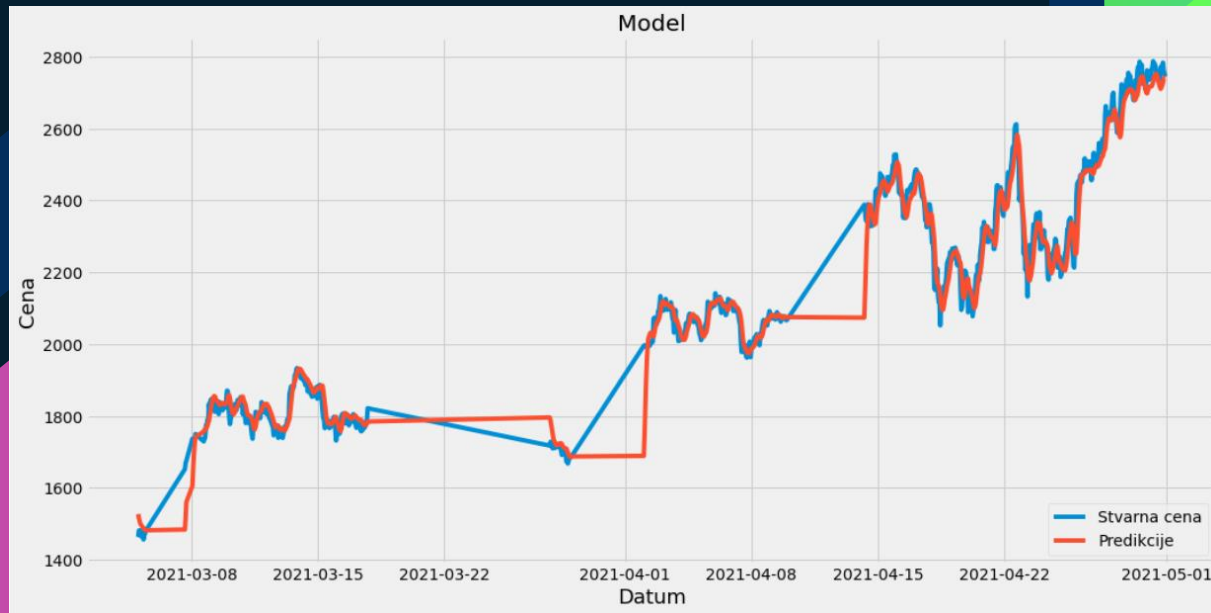
# Графички приказ предвиђених и стварних цена Биткоина



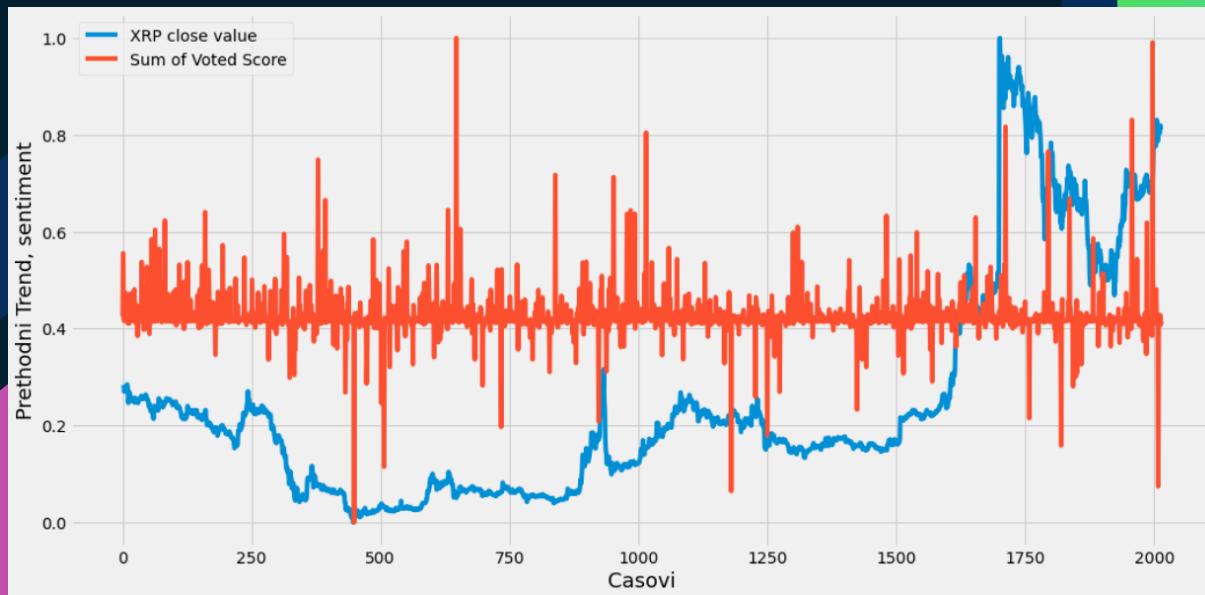
# Корелација цене Етеријума и сентимента



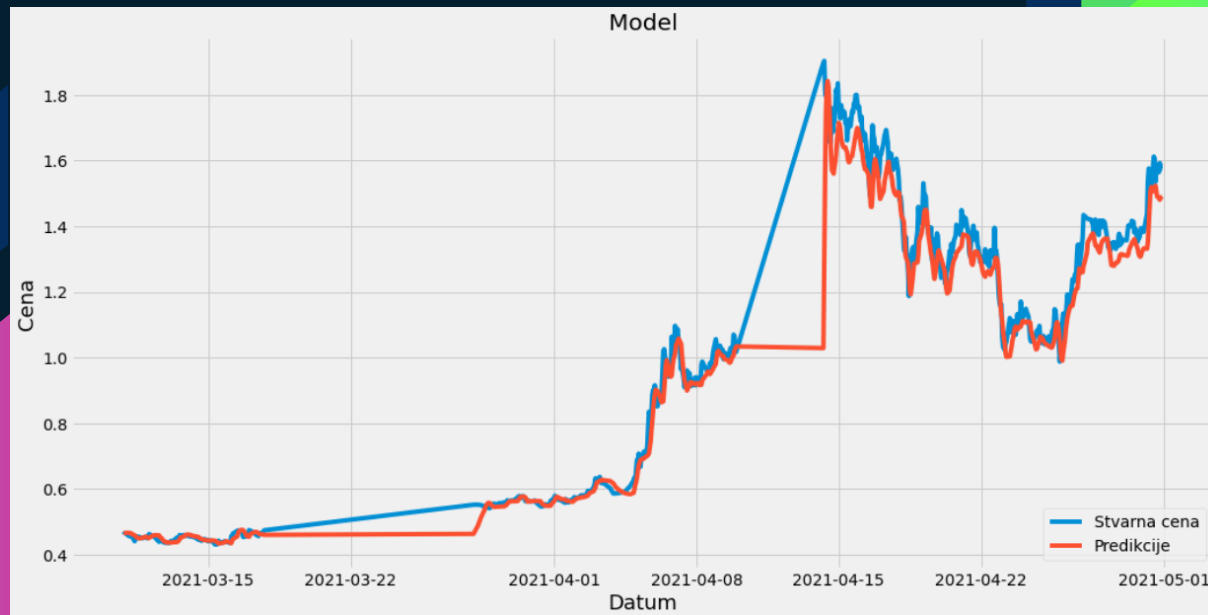
# Графички приказ предвиђених и стварних цена Етеријума



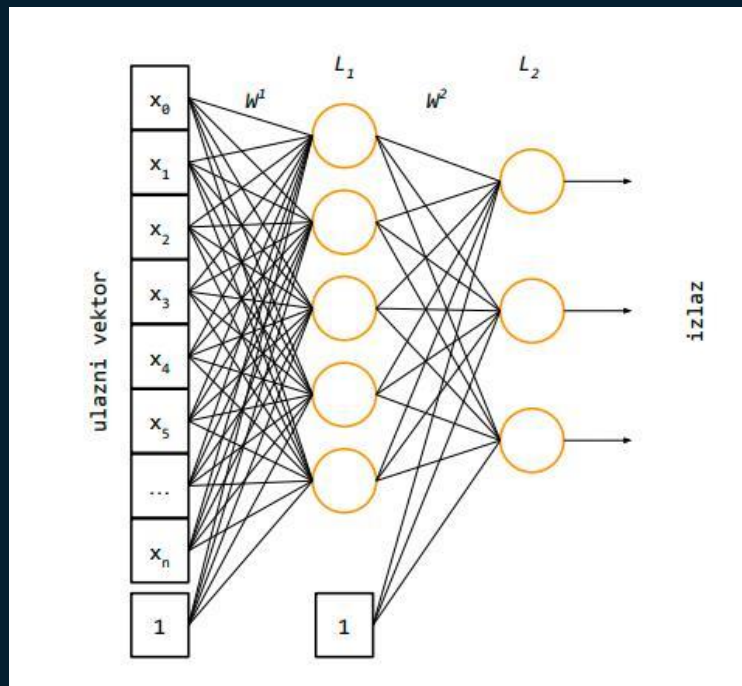
# Корелација цене Рипла и сентимента



# Графички приказ предвиђених и стварних цена Рипла



# Вишеслојни перцептрон





# *Neuroph Framework*

- 3 слоја – улазни, излазни и скривени
- Сигмоидна активациона функција
- 10 неурона у скривеном слоју
- *Momentum Backpropagation Learning Rule*
- *Accuracy*
- *Max Iterations – 25000*
- *Learning Rate = 0.1*
- *Max Error = 0.01*
- *Momentum Value = 0.6*

# ***Neuroph Framework***

- *Binary Classifier Evaluator*
- *Confusion Matrix*
- *Classification Metrics*

# Резултати

	<b>BTC</b>	<b>ETH</b>	<b>XRP</b>
<b>Тачност</b>	<b>0.59</b>	<b>0.56</b>	<b>0.49</b>

Мала тачност?

# Унакрсна корелација

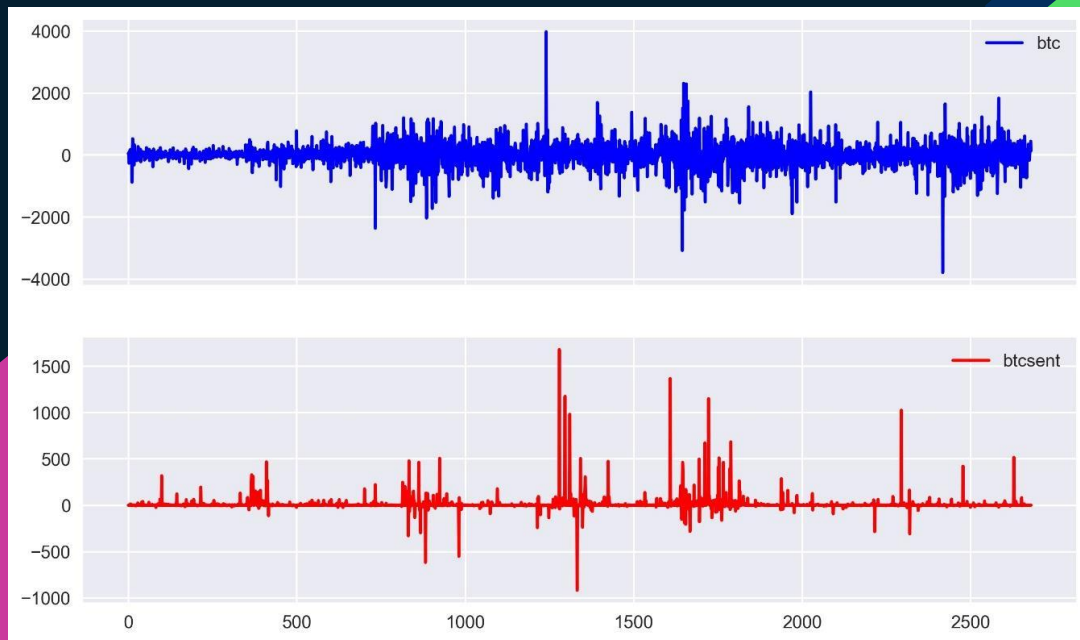
The slide features a dark blue background with abstract, colorful geometric shapes in the corners. On the left, there are overlapping shapes in shades of green, blue, orange, and pink. On the right, there are shapes in shades of green, blue, pink, and orange. The central text is white and reads: 

Промена цене

*лаг?*

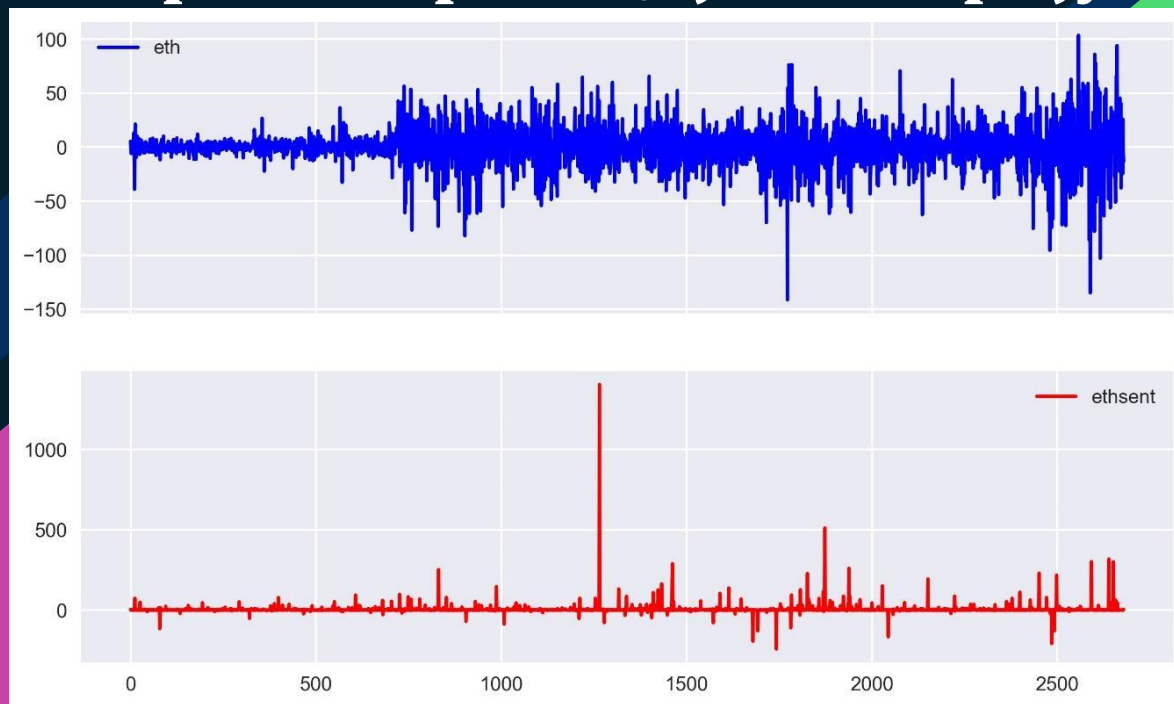
Вредност сентимента

# Унакрсна корелација - Биткоин



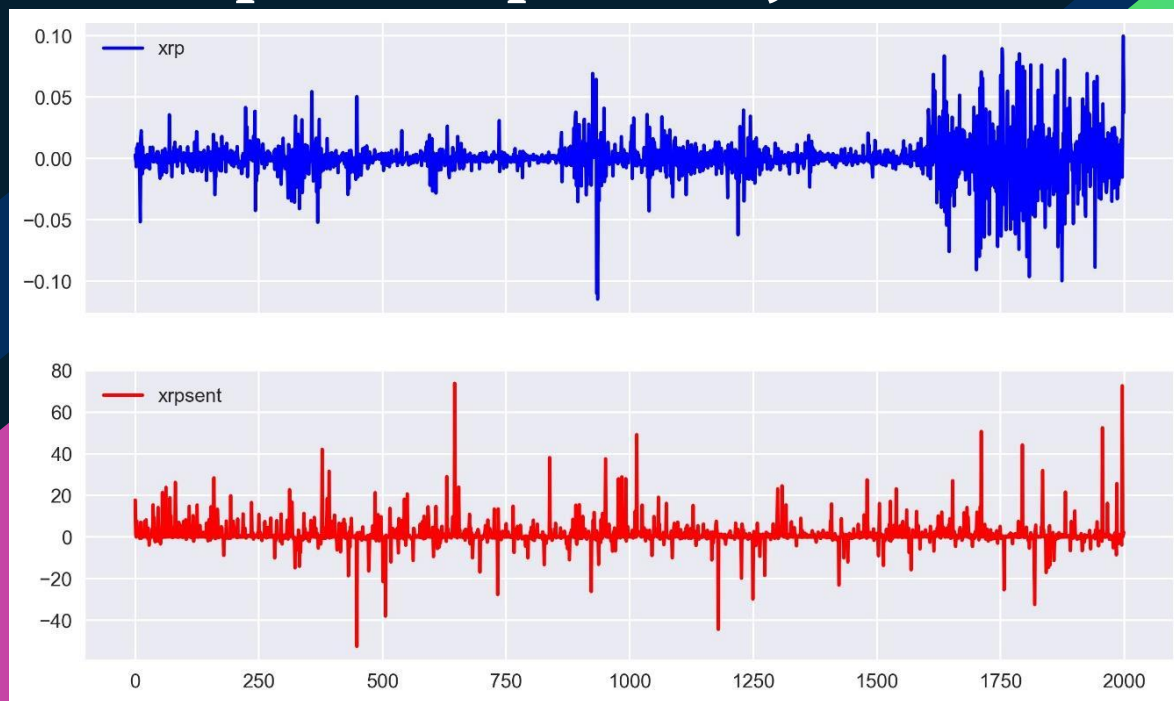
**Лаг 66 – корелација 0.07**

# Унакрсна корелација - Етеријум



**Лаг -1336 – корелација 0.05**

# Унакрсна корелација - Рипл



**Лаг 786 – корелација 0.06**





# Закључак?

# Хвала!

