

## گزارش بخش هوش مصنوعی پروژک های شرکت نورافین

نگین مریوانی

1401/2/18

### پروژک 1

بررسی نرخ بازدهی روزانه بیت کوین با گرفتن داده های candlestick روزانه بیت کوین از yahooFinancials API انجام شده و اختلاف های چشمگیری در قیمت شروع و پایان در بعضی از روزها مشاهده میشود [نمودار 1] که باعث نوسانات زیاد در بازده روزانه شده است [نمودار 2] و با بررسی نمودار 1 میتوان به non-stationary بودن سری زمانی بیت کوین پی برد .

### Bitcoin/USD Chart



Figure1

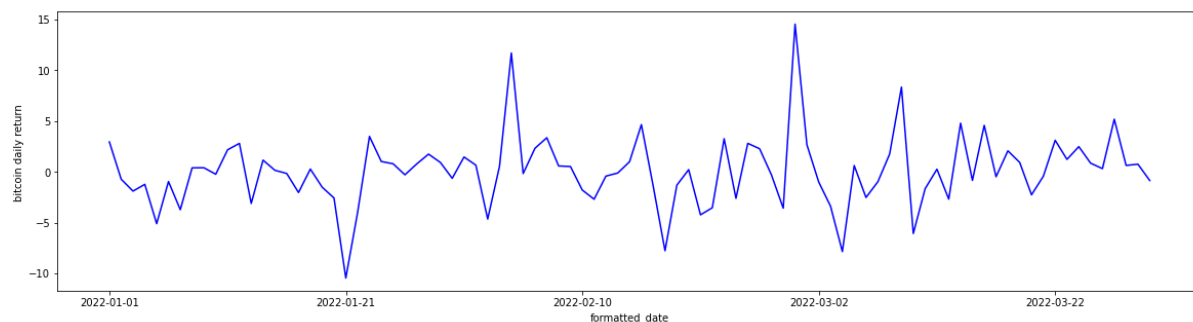


Figure2

برای طبقه بندی نرخ بازده بیتکوین در سه خوشه از دو روش اصلی خوشه بندی مبتنی بر فاصله (kmean) [نمودار 2، 3 و 4] و مبتنی بر توزیع داده (DBSCAN) [نمودار 5] استفاده شده است ، همانطور که دیده میشود الگوریتم kmean در حالت 3 خوشه به صورت تعداد زیادی از داده ها را به کلاس نرخ بازدهی کم و حدود صفر نسبت داده است ولی در حالت 5 خوشه بازده های نزدیک به صفر و بازده های نزدیک به 5 و منها 5 درصد را در خوشه های جداگانه ای دسته بندی کرده و در حالت 8 خوشه میتوان گفت خوشه ها خیلی descriptive نمی باشند. نکته دیگری که میتوان مطرح کرد imbalance بودن داده ها میباشد به نحوی که تعداد روزهای با بازده نزدیک صفر بیشتر از روز های دیگر است که در صورتی که اهمیت روز های دیگر برای ما زیاد باشد میتوانیم با روش های resample این مشکل را حل کنیم .

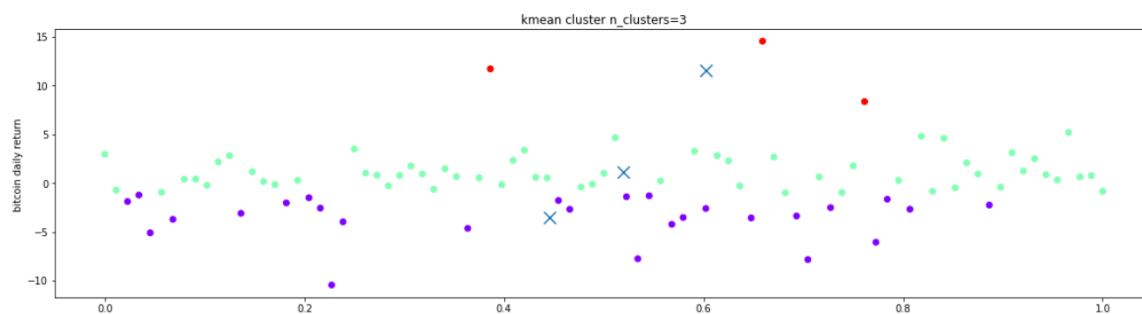


Figure3

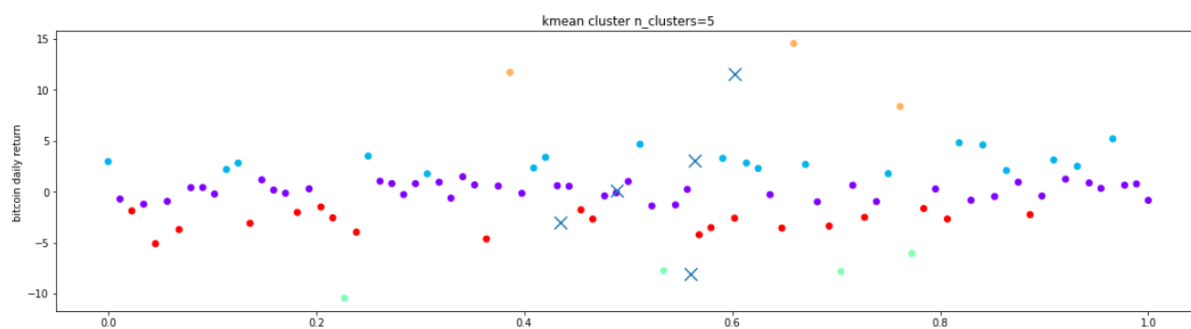


Figure 4

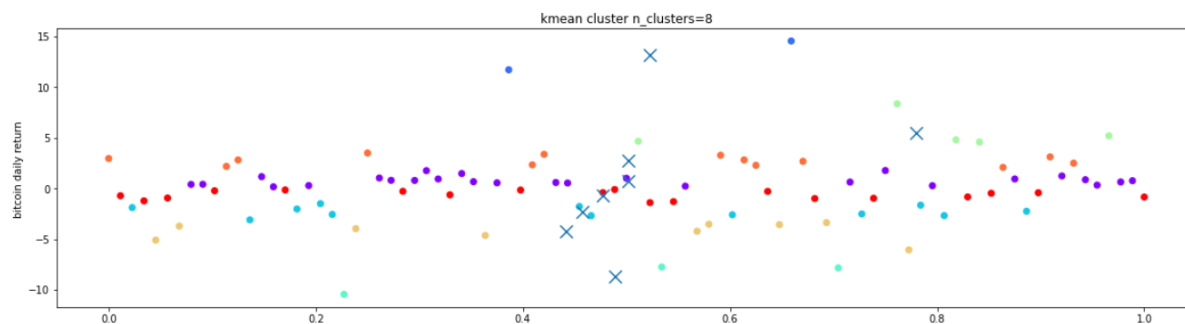


Figure 5

به علت وجود داده های پرت و بعد کم داده ها به نظر می‌رسد استفاده از روش خوشه بندی مبتنی بر توزیع مانند DBSCAN نیز بتواند مفید باشد. اما با وجود هایپرپارامترهای بیشتر در این الگوریتم به دست آوردن خوشه بندی مطلوب سخت‌تر است و میتوان گفت تغییر توزیع نرخ بازده بیت‌کوین در بازه های زمانی متفاوت دقت این مدل را کاهش داده. و شاید تحلیل بازده بیت کوین با مدل های markov switching بهتر خواهد بود.

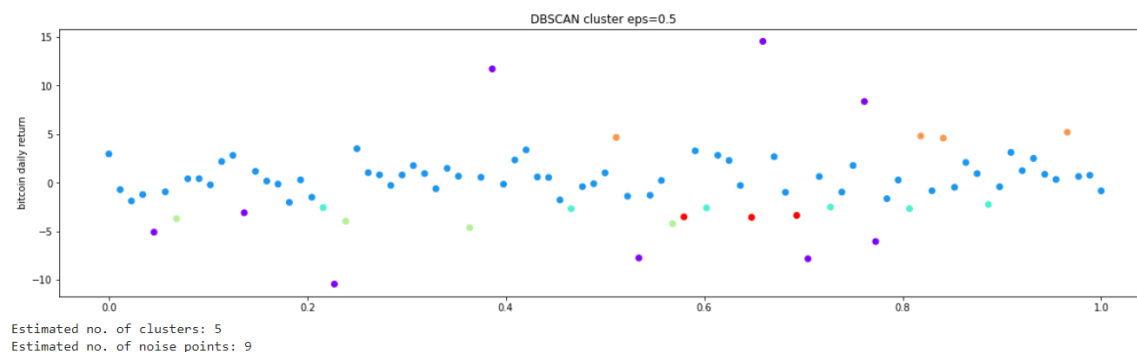


Figure6

## پروژک 2

برای بررسی همبستگی بازده بیت کوین و بازده طلا و نفت خام [نمودارهای 7 و 8] از داده های موجود در yfinance API استفاده کرده و همبستگی این سری های زمانی را در lag 10 مثبت و منفی بررسی شده است.

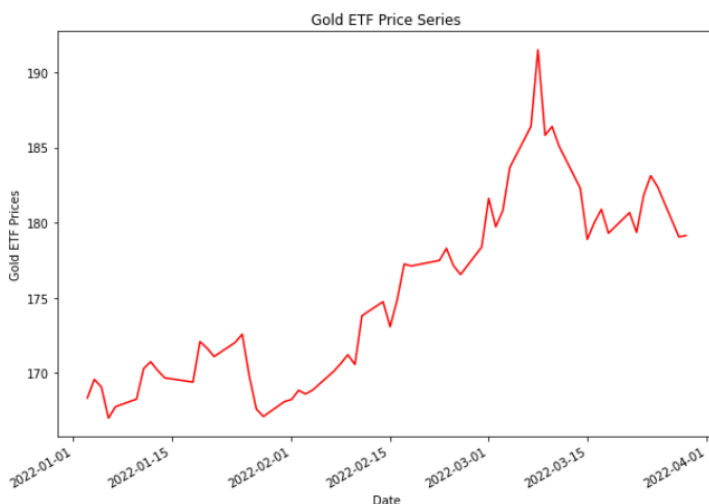


Figure7

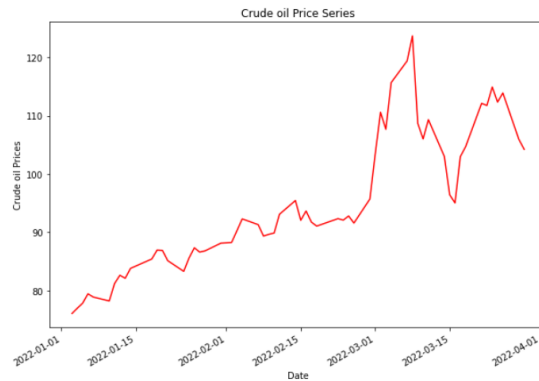


Figure8

در این بخش دو همبستگی pearson (همبستگی خطی) [نمودار 9] و spearman [نمودار 10] بررسی شده است همانطور که دیده میشود بازده بیت کوین همبستگی بیشتری (قدرمطلق همبستگی مورد توجه است) با لگ های منها 5 و مثبت 6 بازده طلا نسبت به لگ صفر ان دارد ( تغییرات در بازار طلا زمان میبرد تا در بازار رمزارز ها اثر بگذارد ) همچنین بازده بیت کوین همبستگی بیشتری با لگ منها 7 نفت خام نسبت به لگ صفر ان دارد . اما در کل میتوان گفت همبستگی بازده بیتکوین و این دو شاخص اقتصادی قابل توجه نمی باشد .

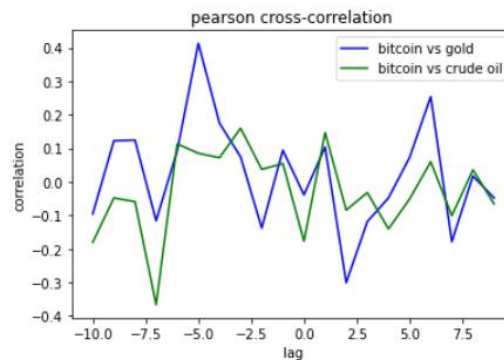


Figure9

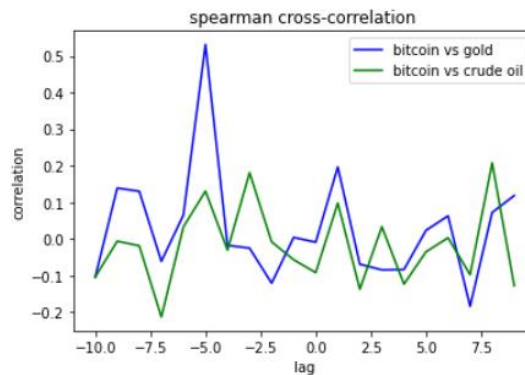


Figure10

### پروژک 3

با توجه به اینکه کلاستر های ساخته شده در پروژک 1 به صورت unsupervised و با دقت کم میباشند در این بخش پیشبینی بازده روزانه بیت کوین با استفاده از بازده های 4 روز گذشته انجام شده است از مدل LSTM به علت توانایی بهتر برای نگهداری حافظه بلند مدت نسبت به مدل های RNN استفاده شده است و پیشبینی مدل برای بخش تست دیتاست در نمودار 11 نمایش داده شده است .

مدل با توجه به استفاده نکردن از هیچ ویژگی دیگری به دقت قابل قبول  $MSE = 1.17$  رسید ( برای بهتر شدن دقت مدل از بازه زمانی بیشتری نسبت به پروژک های قبلی استفاده شده است و ساختار مدل نیز در نمودار 12 بیان شده است )

MSE of LSTM model on test data  
1.1794265053501551

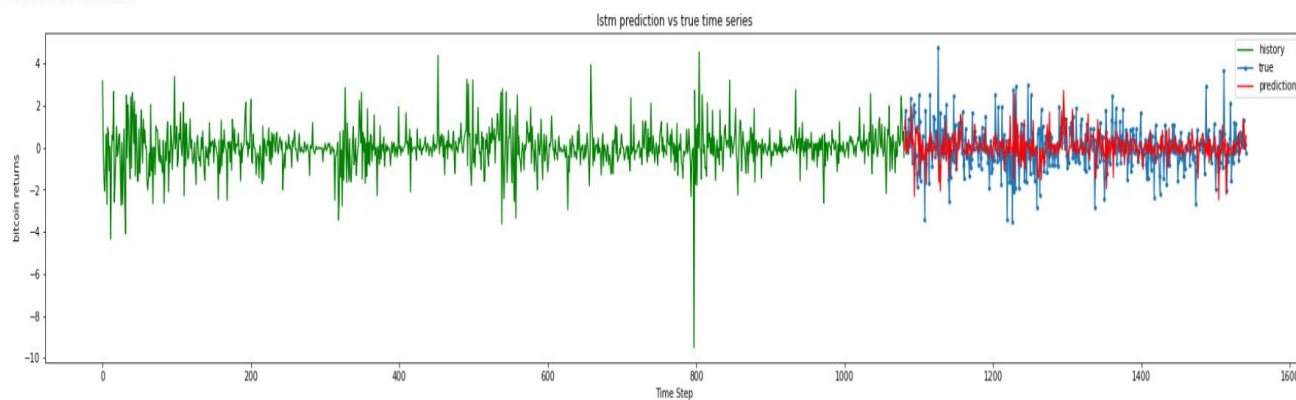


Figure11

Model: "sequential\_21"

Layer (type)	Output Shape	Param #
lstm_21 (LSTM)	(None, 128)	66560
dropout_3 (Dropout)	(None, 128)	0
dense_33 (Dense)	(None, 10)	1290
dense_34 (Dense)	(None, 1)	11

=====  
Total params: 67,861  
Trainable params: 67,861  
Non-trainable params: 0

Figure12