

Research article

Prediction of bitcoin stock price using feature subset optimization

Shiva Vafadar 810899074

Machine Learning course - spring 1403

Professor Mirzaei

در پرتو نوسانات اخیر ارزش ارزهای رمزنگاری شده، بیت‌کوین به تدریج به عنوان یک وسیله سرمایه‌گذاری شناخته شده در حال رشد است. با توجه به نوسانات طبیعی بازار، پیش‌بینی دقیق ارزشی بیت‌کوین برای تصمیم‌گیری‌های سرمایه‌گذاری مطلوب بسیار حیاتی است. به طور ویژه، تحقیقات قبلی از روش‌های یادگیری ماشین برای افزایش دقت پیش‌بینی قیمت بیت‌کوین استفاده کردند. با این حال، تعداد کمی از مطالعات به بررسی پتانسیل استفاده از روش‌های متنوع مدل‌سازی برای نمونه‌برداری با فرمت‌های داده و ویژگی‌های بعدی مختلف پرداخته‌اند. این مطالعه به دنبال شناسایی زیرمجموعه ویژگی داخلی است که باعث بهبود بالاترین بازده پیش‌بینی قیمت بیت‌کوین می‌شود. به طور خاص، ویژگی‌های داخلی بیت‌کوین به چهار گروه دسته‌بندی شدند: داده‌های ارز، جزئیات بلوک، اطلاعات ماینینگ و پیچیدگی شبکه. سپس، یک شبکه عصبی مصنوعی با طول کوتاه مدت (LSTM) برای پیش‌بینی قیمت نهایی بیت‌کوین روز بعد، با استفاده از دسته‌بندی‌های مختلفی از زیرمجموعه ویژگی استفاده شد. مدل با استفاده از دو سال و نیم داده تاریخی هر ویژگی آموزش داده شد. یافته‌ها نشان دادند که نرخ خطای میانگین مطلق 6.38% در هنگام مدل‌سازی با ویژگی‌های دسته جزئیات بلوک به دست آمد. این عملکرد به طور اصلی از ارتباط مثبت بین قیمت بیت‌کوین و کمترین ابهام این زیرمجموعه داده برمی‌آید. نتایج آزمایشی نشان دادند که نسبت به سایر زیرمجموعه‌های ویژگی مورد بررسی، دسته‌بندی ویژگی‌های جزئیات بلوک پیش‌بینی دقیق‌تری از قیمت بیت‌کوین ارائه می‌دهد که پایه‌گذار تحقیقات آینده در این زمینه می‌شود.

پیش‌بینی قیمت بیت‌کوین، سهمیه‌های انحرافی و غیرپایدار معمولاً در سری‌های زمانی بازار بیت‌کوین با فرم غیرخطی و غیرایستا مواجه است. به پاسخ به دو سوال کلیدی تحقیق، بحث‌های گرمی به وجود آمده است: دقیقاً چه چیزی باعث ارزش بیت‌کوین می‌شود؟ چه عواملی بر ارزش بیت‌کوین تأثیر می‌گذارند؟ به طور جالب، ارزش بیت‌کوین نمایانگر اعتقاد سرمایه‌گذاران به ارزهای رمزنگاری شده در زمینه نوآوری مالی است. به همین دلیل، تعداد زیادی از مطالعات بر فاکتورهای موثر یا شکل‌دهنده قیمت بیت‌کوین تمرکز داشته‌اند. علاوه بر این، نوسانات قیمتی ناشی از نوسانات ذاتی بیت‌کوین از زمان ظهور این ارز موجب نگرانی سرمایه‌گذاران شده است. پیش‌بینی تغییرات قیمت بیت‌کوین از اهمیت برجسته‌ای برخوردار است. به همین منظور، پیش‌بینی بازارهای سهام در چند دهه گذشته به دلیل دسترسی به داده‌های روزانه و فرکانس بالا افزایش یافته است. با این حال، تحقیقات کمی برای پیش‌بینی قیمت بیت‌کوین وجود داشته است. تحقیقات قبلی دو روش برای پیش‌بینی قیمت بیت‌کوین به کار بردند: تجزیه و تحلیل تجربی و ارزیابی الگوریتم‌های قدرتمند یادگیری ماشین. به ویژه، الگوریتم‌های یادگیری ماشین به طور گسترده در بخش‌های مختلفی مانند تولید و مالی استفاده شده‌اند. این الگوریتم‌ها و مدل‌ها، با تولید پیش‌بینی‌ها بر اساس داده‌های آموزشی، می‌توانند با تجزیه و تحلیل جزئیات رویدادهای گذشته، توسعه یابند. با توجه به اینکه قوانین تجارت T+0 مایعیت و نوسان را افزایش می‌دهند، می‌توان الگوریتم‌های مشابهی برای بازار بیت‌کوین و بالقوه تمام منظرهای مالی توسعه داد. بیت‌کوین، یک ارز دیجیتال همتای به همدی است که بر روی یک شبکه از همتاها کار می‌کند. بیت‌کوین امکان انجام تراکنش‌های امن و شفاف را بدون وابستگی به موسسات قانونی سنتی فراهم می‌کند. علاوه بر این، بر روی یک سیستم بسیار امن عمل می‌کند، زیرا تمام تراکنش‌ها عمومی هستند و توسط گره‌های شبکه تأیید می‌شوند. در آوریل 2021، بیت‌کوین به بالاترین ارزش تاریخ خود رسید و با قیمت 63,503.46 دلار بسته شد، که نشان‌دهنده یک افزایش تقریباً 927% در یک سال است. ناشناس بودن بیت‌کوین به همراه افزایش سریع آن در یک بازه زمانی کوتاه، فرصت‌هایی برای سرمایه‌گذاران فراهم می‌کند که منجر به برجسته‌سازی پیش‌بینی قیمت بیت‌کوین شده است. علاوه بر این، توسعه روشی برای پیش‌بینی دقیق قیمت بیت‌کوین با استفاده از الگوریتم‌های یادگیری ماشین امر ضروری است. از آنجا که بیت‌کوین دارای فصلی نمی‌باشد، مدل‌های یادگیری ماشین همچنان مورد استفاده و ارزش قرار می‌گیرند. به طور جالب، روش‌های یادگیری ماشین مختلف اساسی مانند شبکه‌های عصبی بازگشتی (RNN)، حافظه کوتاه مدت بلند (LSTM)، ماشین‌های بردار پشتیبان (SVM) و مدل‌های جنگل تصادفی (RF) در تحقیقات قبلی به کار رفته‌اند. با این حال، تحقیقات گذشته اغلب شامل وارد کردن بی‌رویه داده‌ها به مدل‌ها بوده است که از فرکانس داده یا اندازه نمونه صرف‌نظر می‌کند. علاوه بر این، مجموعه‌های داده با فرکانس‌های مختلف ساختارهای متفاوتی دارند و بنابراین، استفاده بی‌تفکر از روش‌های یادگیری ماشین ممکن است منجر به خطاهایی مانند بیش‌باری شود. در عین حال، ویژگی‌های وارد شده به یک شبکه عصبی محدودیت دقت آن را می‌کنند. بنابراین، دقت مدل با انتخاب یک زیرمجموعه بهینه از ویژگی‌های ورودی افزایش می‌یابد، زیرا اطلاعات اضافی وارد ممکن است نویز به سیستم برآورد.

به دلیل نوسانات بیت‌کوین، تعیین اینکه کدام ویژگی‌ها در زمان پیش‌بینی استفاده شود چالشی است. علاوه بر این، ارزش بیت‌کوین به همان شکلی که ارزش‌های سهام شرکت‌ها را تحت تأثیر قرار می‌دهد، تحت تأثیر قرار نمی‌گیرد؛ بنابراین، رویکرد پیش‌بینی باید متفاوت باشد. از این رو، این تحقیق به دنبال بدست آوردن بینش‌هایی برای شناسایی که ویژگی‌های داخلی بیت‌کوین باید برای پیش‌بینی ارزش بیت‌کوین استفاده شوند. بطور دقیق، ویژگی‌های بیت‌کوین به چهار دسته تقسیم شده‌اند: داده‌های ارز، جزئیات بلوک، اطلاعات ماینینگ و پیچیدگی شبکه، که در شکل ۱ نمایش داده شده‌اند. درک ارتباط بین ویژگی‌های شناسایی شده و قیمت نهایی روزانه می‌تواند برای تحقیقات آینده مفید باشد. در این تحقیق، یک شبکه عصبی LSTM برای پیش‌بینی قیمت نهایی بیت‌کوین روز بعد استفاده شده است. به طور خاص، مدل با استفاده از زیرمجموعه‌های ویژگی آموزش دیده و دقت پیش‌بینی حاصل مقایسه شده‌است. این تحقیق با توسعه یک تکنیک مدل‌سازی پیش‌بینی نوآورانه برای پیش‌بینی سری‌های زمانی، ترکیب بهینه‌سازی و چارچوب بازگشتی در یک محیط گرانولی، به توسعه یک روش پیش‌بینی دقیق قیمت بیت‌کوین کمک می‌کند.

پیش‌بینی سری‌های زمانی بیت‌کوین

بیت‌کوین به عنوان قدیمی‌ترین و شناخته‌شده‌ترین ارز دیجیتال، اولین بار در سال ۲۰۰۹ توسط فرد یا گروهی ناشناس با نام ساتوشی ناکاموتو به صورت متن‌باز منتشر شد. بیت‌کوین به عنوان یک وسیله تبادل دیجیتال غیرمتمرکز عمل می‌کند، با تراکنش‌هایی که در یک دفترکل عمومی توزیع شده (بلاکچین) بدون نیاز به یک مرجع معتبر ثبت یا واسطه مرکزی تایید و ثبت می‌شوند. بلوک‌های تراکنش حاوی یک هش رمزنگاری شده SHA-256 از بلوک‌های تراکنش قبلی هستند و به این ترتیب به صورت "زنجیره‌ای" به هم متصل شده‌اند و به عنوان یک رکورد غیرقابل تغییر از تمام تراکنش‌هایی که تا به حال انجام شده‌اند، خدمت می‌کنند. مانند هر ارز/کالای دیگری در بازار، تجارت بیت‌کوین و ابزارهای مالی به زودی پس از پذیرش عمومی بیت‌کوین به وجود آمدند و همچنان رشد می‌کنند. اگر نمی‌دانید بیت‌کوین چیست، اطلاعاتی درباره بیت‌کوین را اینجا دریافت کنید. این کرنل به دو بخش تقسیم شده است:

کاوش داده‌ها

تحلیل سری‌های زمانی

و بیشتر برای پیش‌بینی سری‌های زمانی:

پیش‌بینی سری‌های زمانی با LSTM

پیش‌بینی سری‌های زمانی با XGBoost

پیش‌بینی سری‌های زمانی با Facebook Prophet

پیش‌بینی سری‌های زمانی با ARIMA

کاوش داده‌ها

در این بخش، فقط به کاوش داده‌ها یعنی قیمت‌های تاریخی بیت‌کوین می‌پردازیم و سعی می‌کنیم به برخی از بینش‌ها دست پیدا کنیم. ما از مجموعه داده‌های Coinbase استفاده

خواهیم کرد زیرا یکی از پراستفاده‌ترین صرافی‌ها/کیف پول‌های بیت‌کوین در جهان است.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2016354 entries, 0 to 2016353
Data columns (total 8 columns):
Timestamp                datetime64[ns, UTC]
Open                     float64
High                     float64
Low                      float64
Close                    float64
Volume_(BTC)             float64
Volume_(Currency)       float64
Weighted_Price           float64
dtypes: datetime64[ns, UTC](1), float64(7)
memory usage: 123.1 MB
```

پیش‌پردازش کوچکی مورد نیاز است، جایگزینی مقادیر NaN با صفرها و داده‌های قبلی.

In [5]:

```
# First thing is to fix the data for bars/candles where there are no trades.
# Volume/trades are a single event so fill na's with zeroes for relevant fields...
data['Volume_(BTC)'].fillna(value=0, inplace=True)
data['Volume_(Currency)'].fillna(value=0, inplace=True)
data['Weighted_Price'].fillna(value=0, inplace=True)

# next we need to fix the OHLC (open high low close) data which is a continuous timeseries so
# lets fill forwards those values...
data['Open'].fillna(method='ffill', inplace=True)
data['High'].fillna(method='ffill', inplace=True)
data['Low'].fillna(method='ffill', inplace=True)
data['Close'].fillna(method='ffill', inplace=True)

data.head()
```

Out[5]:

	Timestamp	Open	High	Low	Close	Volume_(BTC)	Volume_(Currency)	Weighted_Price
0	2014-12-01 05:33:00+00:00	300.0	300.0	300.0	300.0	0.01	3.0	300.0
1	2014-12-01 05:34:00+00:00	300.0	300.0	300.0	300.0	0.00	0.0	0.0
2	2014-12-01 05:35:00+00:00	300.0	300.0	300.0	300.0	0.00	0.0	0.0
3	2014-12-01 05:36:00+00:00	300.0	300.0	300.0	300.0	0.00	0.0	0.0
4	2014-12-01 05:37:00+00:00	300.0	300.0	300.0	300.0	0.00	0.0	0.0

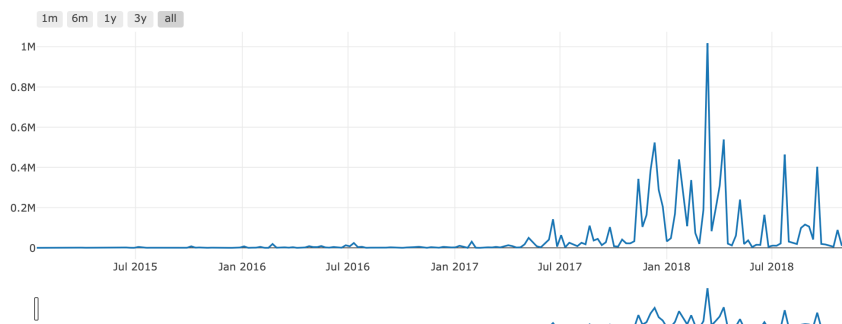
بیا ببینیم قیمت‌های تاریخی بیت‌کوین را در بازه زمانی ۲۰۱۵ تا ۲۰۱۸ بصورت بصری نمایش دهیم.

Historical Bitcoin Prices (2015-2018) with the Slider

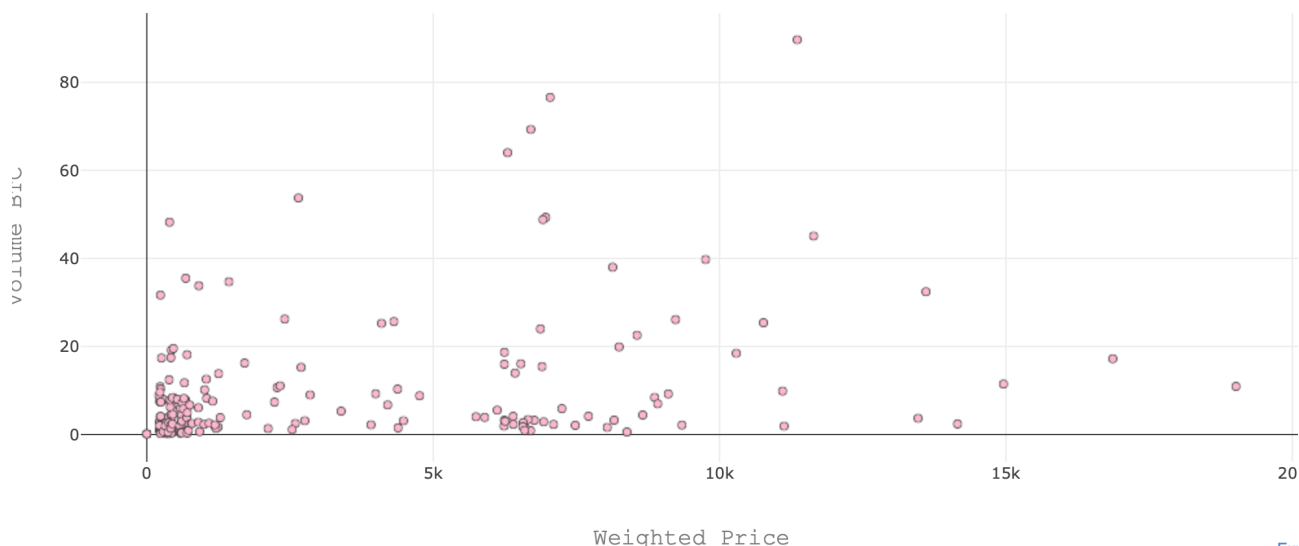


بیا ببینیم حجم بازار تاریخی بیت‌کوین را در بازه زمانی ۲۰۱۵ تا ۲۰۱۸ بصورت بصری نمایش دهیم.

Historical Bitcoin Volume (USD) (2015-2018) with the slider



BTC Volume v/s USD



تاریخچه قیمت‌های بیت‌کوین

۱۸ آگوست ۲۰۰۸، دامنه bitcoin.org ثبت شد.

۶ نوامبر ۲۰۱۰، سرمایه بیت‌کوین به ۱ میلیون دلار آمریکا رسید. نرخ تبدیل آن در MtGox به ۰.۵۰ دلار آمریکا به ازای هر BTC رسید.

۲ ژوئن ۲۰۱۱، نرخ دلار به بیت‌کوین ۱۰ دلار به ازای هر سکه بود. به مدت ۶ روز، ارزش بیت‌کوین در MtGox به ۳۱.۹۱ دلار ثابت ماند.

۲۸ فوریه ۲۰۱۳، نرخ تبدیل بیت‌کوین برای اولین بار در ۶۰۱ روز گذشته از ۳۱.۹۱ دلار گذشت.

۱ آوریل ۲۰۱۳، نرخ تبدیل بیت‌کوین به ۱۰۰ دلار به ازای ۱ BTC رسید.

ژانویه ۲۰۱۵، Coinbase در دور سری C جمع‌آوری سرمایه ۷۵ میلیون دلار آمریکا جذب کرد، که رکورد قبلی برای یک شرکت بیت‌کوین را شکست.

فوریه ۲۰۱۵، قیمت بیت‌کوین به ۲۶۲ دلار آمریکا رسید.

ژانویه ۲۰۱۷، پس از رالی بیشتر نیمه دوم سال ۲۰۱۶، بیت‌کوین برای اولین بار در ۳ سال به مرز ۱۰۰۰ دلار آمریکا رسید.

۱۲ ژوئن ۲۰۱۷، نرخ تبدیل بیت‌کوین از ۳۰۰۰ دلار آمریکا به ازای هر BTC گذشت.

۲۹ نوامبر ۲۰۱۷، قیمت بیت‌کوین از ۱۰,۰۰۰ دلار آمریکا گذشت.

۱۸ دسامبر ۲۰۱۷، بیت‌کوین به بالاترین رکورد خود رسید، اما به ۲۰,۰۰۰ دلار آمریکا نرسید.

۲۸ دسامبر ۲۰۱۷، پس از اعلام اقدامات اضافی کره جنوبی برای تنظیم تجارت بیت‌کوین، از جمله احتمال بسته شدن صرافی‌ها، قیمت بیت‌کوین کاهش یافت، در میان حرکات نوسانی در سومین بازار بزرگ ارزهای دیجیتال جهان.

۳۱ اکتبر ۲۰۱۸، ۶,۳۰۰ دلار آمریکا، در دهمین سالگرد بیت‌کوین، قیمت به طور پایدار بالای ۶,۰۰۰ دلار آمریکا باقی ماند، در دوری از نوسان تاریخی کم.

پیش‌بینی سری‌های زمانی

داده‌های سری زمانی داده‌های آزمایشی هستند که در نقاط مختلف زمانی (معمولاً به صورت یکنواخت، مانند یک بار در روز یا یک بار در ساعت یا یک بار در دقیقه) مشاهده شده‌اند. به عنوان مثال، داده‌های فروش بلیط هواپیما در هر روز یک سری زمانی است. اما، صرفاً به دلیل اینکه یک سری از وقایع یک عنصر زمانی دارند، به طور خودکار سری زمانی نمی‌شوند، مانند تاریخ‌های وقوع حوادث بزرگ هوایی که به صورت تصادفی فاصله دارند و سری زمانی نیستند. این نوع فرآیندهای تصادفی به عنوان فرآیند نقطه‌ای شناخته می‌شوند.

سری‌های زمانی چند ویژگی کلیدی دارند مانند روند، فصلی بودن و نویز. پیش‌بینی فرآیند پیش‌بینی آینده، بر اساس داده‌های گذشته و حال است.

در این کرنل، ما تلاش می‌کنیم تا تحلیل سری‌های زمانی را بر روی داده‌های تاریخی قیمت بیت‌کوین انجام دهیم. می‌توانیم به راحتی از بخش کاوش داده‌ها ببینیم که قیمت‌های بیت‌کوین در طول سال‌ها بسیار ناپایدار و ناهمبند بوده است. انجام تحلیل

سری‌های زمانی بر روی چنین داده‌های ناپایدار بسیار سخت است. اما در اینجا ما سعی می‌کنیم مدل‌های مختلف پیش‌بینی سری‌های زمانی را بررسی کنیم. همه مدل‌های استفاده شده در این کرنل بسیار پایه‌ای هستند و امکان استفاده از مدل‌های پیچیده‌تر و با عملکرد بهتر وجود دارد.

پیش‌بینی سری‌های زمانی با LSTM

پیش‌بینی سری‌های زمانی با XGBoost

پیش‌بینی سری‌های زمانی با Facebook Prophet

پیش‌بینی سری‌های زمانی با ARIMA

پیش‌بینی با استفاده از LSTM

در بخش اول، از LSTM (حافظه بلندمدت کوتاه‌مدت) استفاده می‌کنیم. واحدهای LSTM واحدهای یک شبکه عصبی بازگشتی (RNN) هستند. یک RNN متشکل از واحدهای LSTM اغلب به عنوان یک شبکه LSTM (یا فقط LSTM) نامیده می‌شود. یک واحد LSTM معمولی از یک سلول، یک دروازه ورودی، یک دروازه خروجی و یک دروازه فراموشی تشکیل شده است. سلول مقادیر را در بازه‌های زمانی دلخواه به خاطر می‌سپارد و سه دروازه جریان اطلاعات به داخل و خارج از سلول را تنظیم می‌کنند. اگرچه LSTM برای پیش‌بینی بازار ناپایدار مانند بیت‌کوین ایده‌آل نیست، اما همچنان در اینجا یک شانس امتحان می‌کنیم.

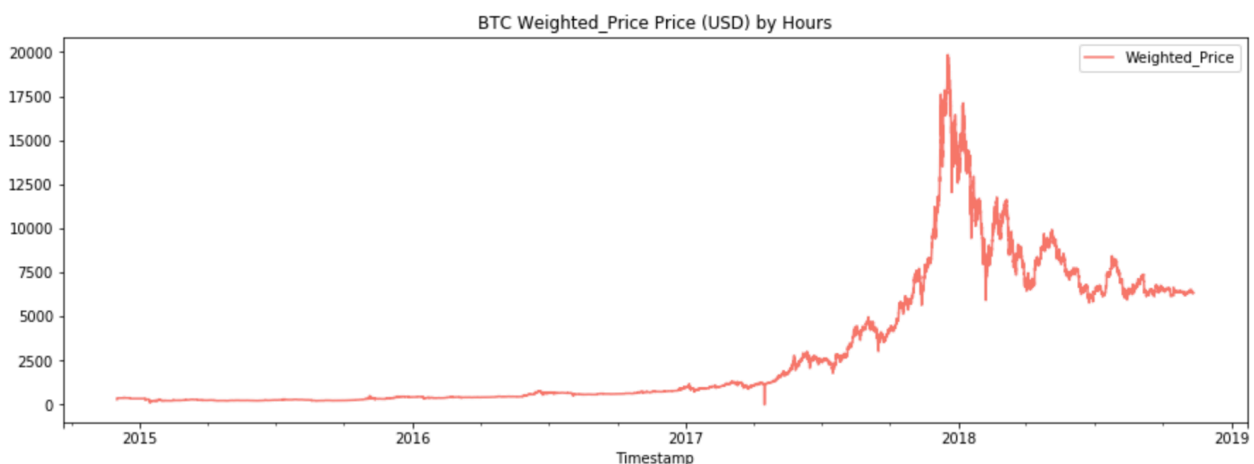
برای تقسیم داده‌ها به آموزش و تست، تاریخ '۲۵-ژوئن-۲۰۱۸' را به عنوان تاریخ تقسیم در نظر می‌گیریم. در دوره ژوئن-ژوئای ۲۰۱۸ افت قابل‌توجهی در قیمت بیت‌کوین وجود داشت. اگر به قیمت‌های تاریخی نگاه کنیم، بازار فصلی از این تاریخ شروع به افزایش کرد پس از رسیدن به کمترین میزان، اگرچه قیمت در ۲۹ ژوئن ۲۰۱۸ به پایین‌ترین مقدار یعنی ۵۹۷۲ دلار رسید. پس از رسیدن به مرز تاریخی ۲۰K در ۱۸ دسامبر، چندین افت رخ داد و هر بار قیمت بازار تصحیح شد.

In [11]:

```
# split data
split_date = '25-Jun-2018'
data_train = data.loc[data.index <= split_date].copy()
data_test = data.loc[data.index > split_date].copy()
```

In [12]:

```
# Data preprocess
training_set = data_train.values
training_set = np.reshape(training_set, (len(training_set), 1))
from sklearn.preprocessing import MinMaxScaler
sc = MinMaxScaler()
training_set = sc.fit_transform(training_set)
X_train = training_set[0:len(training_set)-1]
y_train = training_set[1:len(training_set)]
X_train = np.reshape(X_train, (len(X_train), 1, 1))
```





در اینجا از مدل Vanilla LSTM برای پیش‌بینی استفاده خواهیم کرد. این مدل بر روی داده‌های قبل از ۲۵ ژوئن ۲۰۱۸ آموزش دیده است.

```
In [15]: # Importing the Keras libraries and packages
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from keras.layers import Dropout
from keras.layers import Activation

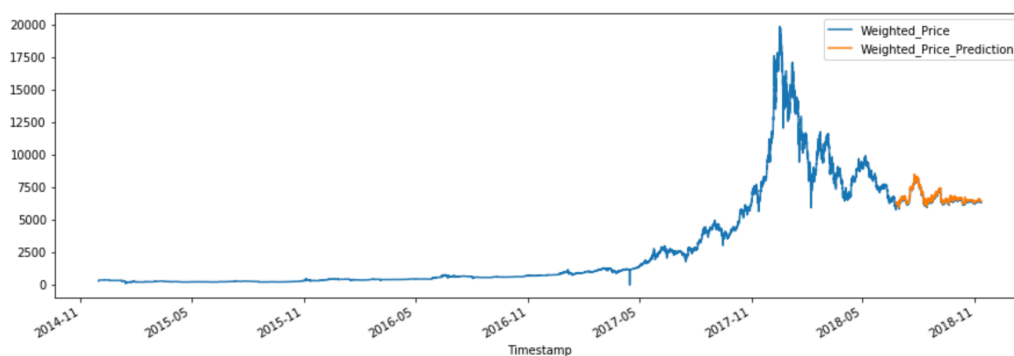
model = Sequential()
model.add(LSTM(128, activation="sigmoid", input_shape=(1,1)))
model.add(Dropout(0.2))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
model.fit(X_train, y_train, epochs=100, batch_size=50, verbose=2)
```

```
In [16]: model.summary()
```

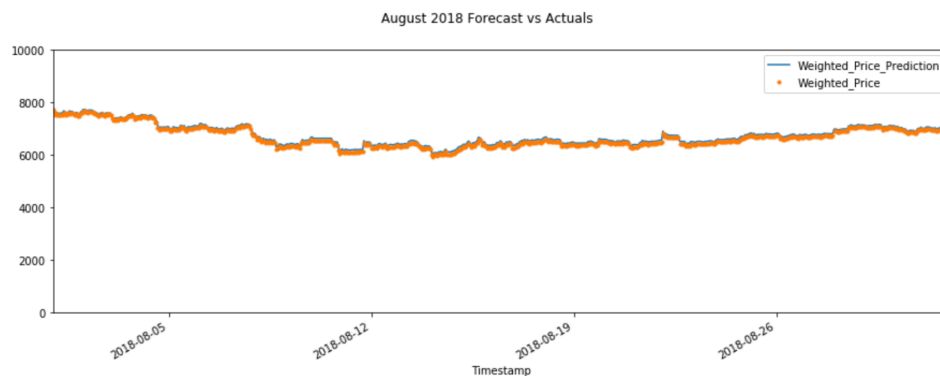
Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 128)	66560
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

Total params: 66,689
 Trainable params: 66,689
 Non-trainable params: 0

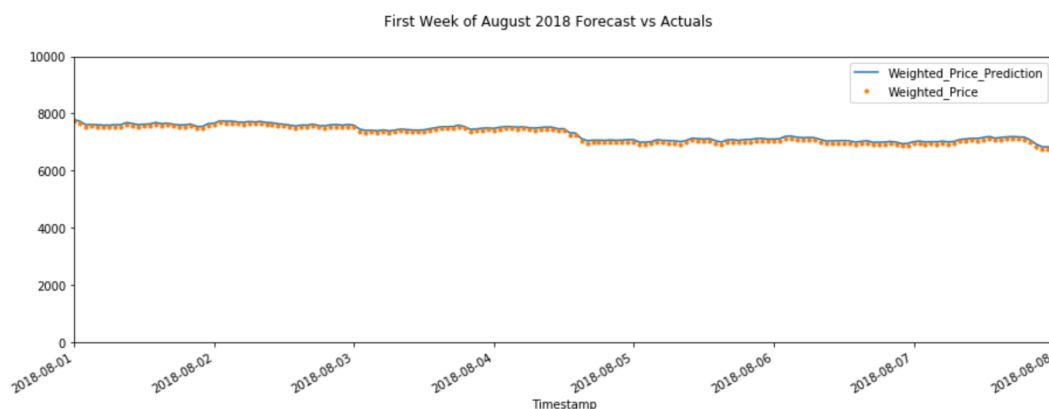
```
In [20]: _ = data_all[['Weighted_Price', 'Weighted_Price_Prediction']].plot(figsize=(15, 5))
```



```
In [21]: # Plot the forecast with the actuals
f, ax = plt.subplots(1)
f.set_figheight(5)
f.set_figwidth(15)
_ = data_all[['Weighted_Price_Prediction', 'Weighted_Price']].plot(ax=ax,
                                                                    style=['-', '.'])
ax.set_xbound(lower='08-01-2018', upper='09-01-2018')
ax.set_ylim(0, 10000)
plot = plt.suptitle('August 2018 Forecast vs Actuals')
```



```
In [22]: # Plot the forecast with the actuals
f, ax = plt.subplots(1)
f.set_figheight(5)
f.set_figwidth(15)
_ = data_all[['Weighted_Price_Prediction', 'Weighted_Price']].plot(ax=ax,
                                                                    style=['-', '.'])
ax.set_xbound(lower='08-01-2018', upper='08-08-2018')
ax.set_ylim(0, 10000)
plot = plt.suptitle('First Week of August 2018 Forecast vs Actuals')
```



خطای میانگین مربعات (MSE) برابر با ۲۹۱۷.۴۰ و خطای میانگین مطلق (MAE) برابر با ۵۳.۸۴ است. می‌توانیم به وضوح ببینیم که مدل به طور کلی بیش‌برازش شده است.

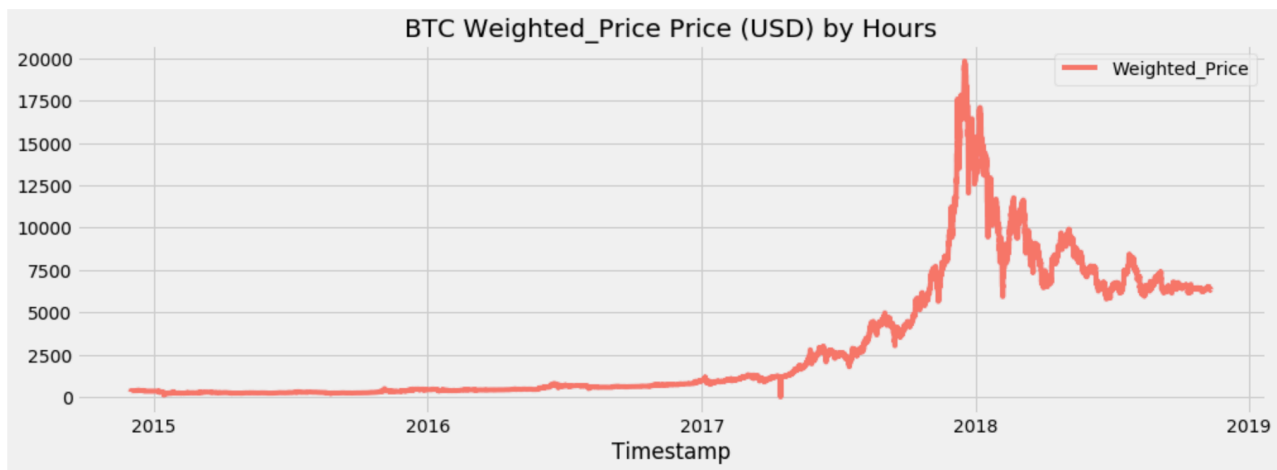
پیش‌بینی سری زمانی با XGBoost

XGBoost یک پیاده‌سازی از درختهای تصمیم تقویت شده با گرادین است که برای سرعت و عملکرد بهینه طراحی شده است. XGBoost ابزاری قدرتمند و چندمنظوره است. بیایید ببینیم که XGBoost چقدر عملکرد خوبی ارائه می‌دهد وقتی از آن برای پیش‌بینی ارزش‌های آینده یک سری زمانی مانند قیمت‌های بیت‌کوین استفاده می‌کنیم.

```
In [25]: import seaborn as sns
import matplotlib.pyplot as plt
from fbprophet import Prophet
from sklearn.metrics import mean_squared_error, mean_absolute_error
plt.style.use('fivethirtyeight')
```

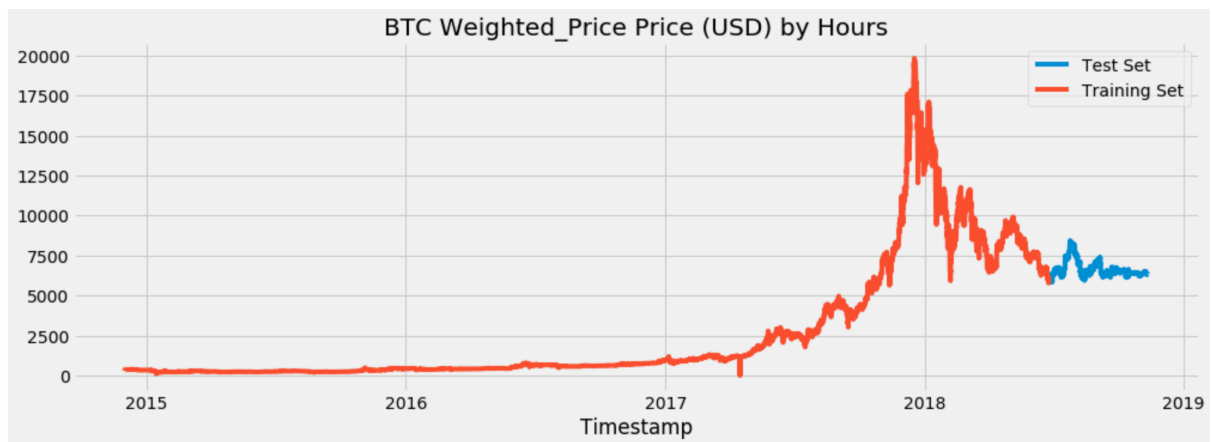
```
In [26]: data = pd.read_csv('../input/coinbaseUSD_1-min_data_2014-12-01_to_2018-11-11.csv', parse_dates=[0], date_parser=dateparse)
data['Timestamp'] = data['Timestamp'].dt.tz_localize(None)
data = data.groupby([pd.Grouper(key='Timestamp', freq='H')]).first().reset_index()
data = data.set_index('Timestamp')
data = data[['Weighted_Price']]
data['Weighted_Price'].fillna(method='ffill', inplace=True)
```

```
In [27]: color_pal = ["#F8766D", "#D32900", "#93AA00", "#00BA38", "#00C19F", "#00B9E3", "#619CFF", "#DB72FB"]
_ = data.plot(style='', figsize=(15,5), color=color_pal[0], title='BTC Weighted_Price Price (USD) by Hours')
```

In [29]:

```
_ = data_test \
    .rename(columns={'Weighted_Price': 'Test Set'}) \
    .join(data_train.rename(columns={'Weighted_Price': 'Training Set'}), how='outer') \
    .plot(figsize=(15,5), title='BTC Weighted_Price Price (USD) by Hours', style='')
```



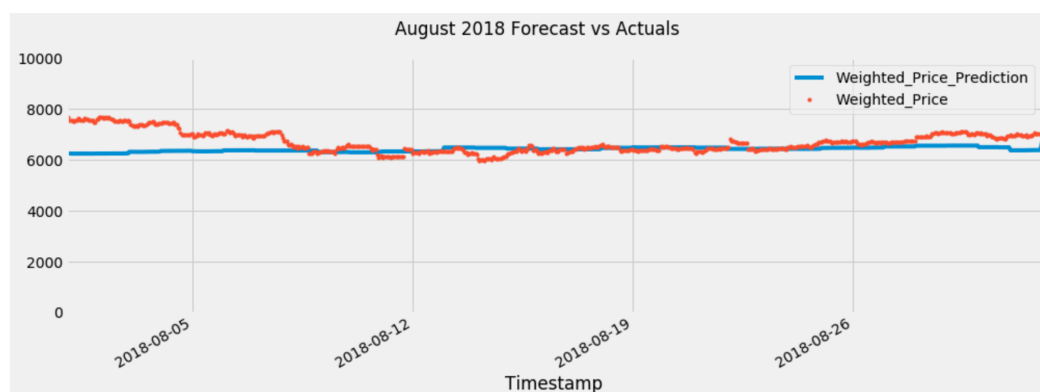
اینجا از یک مدل پایه XGBRegressor استفاده می‌کنیم.

Out[32]:

```
XGBRegressor(alpha=10, base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bytree=0.3, gamma=0, learning_rate=0.1, max_delta_step=0,
              max_depth=5, min_child_weight=10, missing=None, n_estimators=100,
              n_jobs=1, nthread=None, objective='reg:linear', random_state=0,
              reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None,
              silent=True, subsample=1)
```

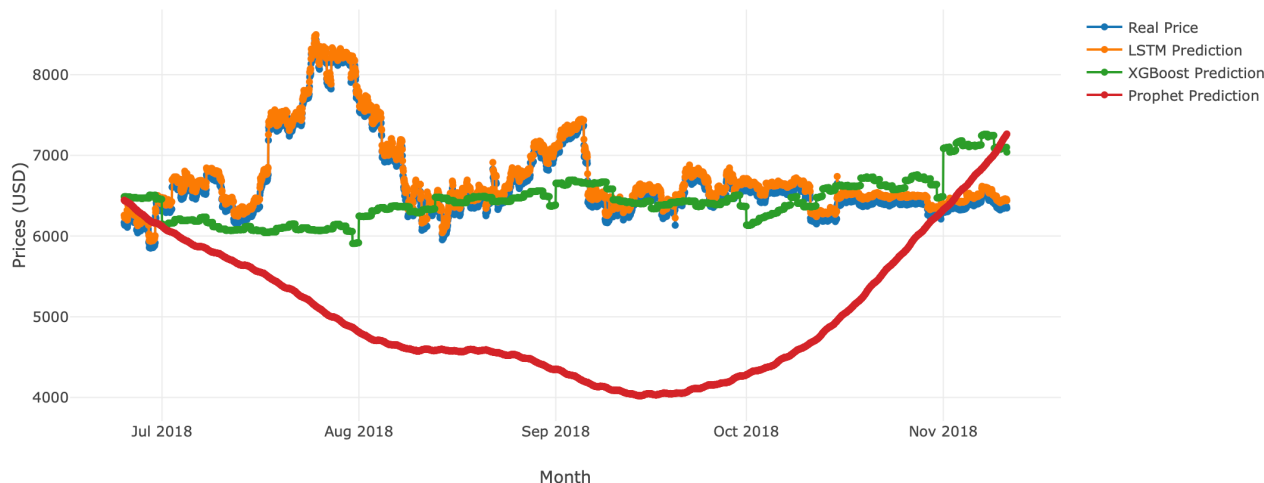
In [36]:

```
# Plot the forecast with the actuals
f, ax = plt.subplots(1)
f.set_figheight(5)
f.set_figwidth(15)
_ = data_all[['Weighted_Price_Prediction', 'Weighted_Price']].plot(ax=ax,
                                                                    style=['-', '.'])
ax.set_xbound(lower='08-01-2018', upper='09-01-2018')
ax.set_ylim(0, 10000)
plot = plt.suptitle('August 2018 Forecast vs Actuals')
```



خطای میانگین مربعات (MSE) برابر با ۴۸۴۴۷۶.۶۹ و خطای میانگین مطلق (MAE) برابر با ۴۷۴.۸۱ است. ال‌اس‌تی‌ام، ایکس‌جی‌بوست و پروفیت - عملکرد آنها چقدر است؟ در اینجا بیاید نتایج پیش‌بینی را برای ال‌اس‌تی‌ام، ایکس‌جی‌بوست و پروفیت در یک نمودار واحد مقایسه و مشاهده کنیم.

Comparison of LSTM, XGBoost and Prophet



می‌توانیم ببینیم که چگونه مدل‌های فوق با داده‌های تاریخی قیمت بیت‌کوین عمل می‌کنند. پیش‌بینی خوبی ارائه نمی‌دهند. قیمت‌های بیت‌کوین بسیار ناپایدار و تصادفی هستند و اغلب تحت تأثیر عوامل خارجی (مانند مقررات رمزارزها، سرمایه‌گذاری‌ها یا شایعات ساده در رسانه‌های اجتماعی) قرار می‌گیرند. برای بهبود و دقیق‌تر کردن عملکرد این مدل‌ها، نیاز به داده‌های اضافی از خبرها یا رسانه‌های اجتماعی داریم.

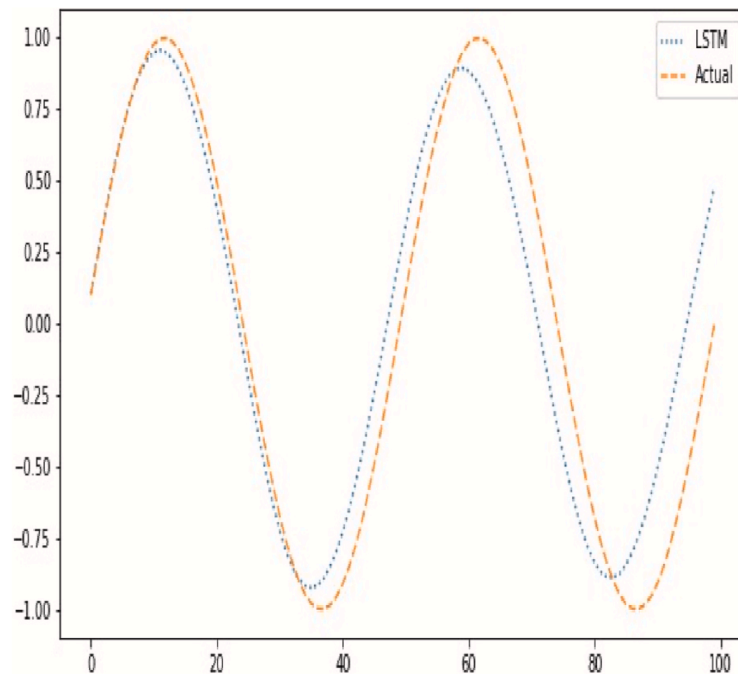
عملکرد تکنیک پیشنهادی با استفاده از سری زمانی اصلی قیمت بیت‌کوین و یک سری زمانی جایگزین به مدت 8 ماه (نوامبر 2020 تا جولای 2021) که به صورت تصادفی از داده‌های اصلی با استفاده از روش مونته کارلو ایجاد شده بود، ارزیابی شد. سری زمانی جایگزین به وضوح بیشترین ناپیش‌بینی و آشوب را نشان می‌دهد. این ارزیابی با استفاده از سری‌های جایگزین برای بررسی قدرت و کارایی روش پیش‌بینی پیشنهادی بسیار حیاتی بود، با حفظ تنظیمات پارامتری به طور مداوم در هر دو مورد.

مدل شامل سه لایه و یک تابع فعال‌سازی خطی بود که طبق پیشنهادات ارجاعات، به عنوان پیکربندی بهینه با در نظر گرفتن تمام ویژگی‌ها ثابت شد. یک dropout لایه‌ای با نرخ 0.2 به هر لایه LSTM اضافه شد تا از بیش‌پرازش جلوگیری کند با کاهش حساسیت به وزن‌های خاص نورون‌ها. بهینه‌ساز Adam با تابع خطای میانگین مربعات (MSE) استفاده شد.

تعداد ایپوک‌ها به 50 محدود شد تا هزینه محاسباتی مدیریت شود و اندازه دسته به 10 تنظیم شد. مدل برای هر مجموعه ویژگی آزمایش شد، با تعداد گرهای تنظیم شده به 32 و 64 و مقدار نگاه به عقب به 30 و 60 روز. مقدار نگاه به عقب تعداد روزهای قبلی ویژگی‌های استفاده شده برای پیش‌بینی قیمت بیت‌کوین در روز بعد را نشان می‌دهد.

علاوه بر این، ارزیابی عملکرد مدل بدون داده بسیار حیاتی است، به ویژه زمانی که بر پیش‌بینی‌های تاریخی برای پیش‌بینی نتایج آینده اعتماد می‌کنیم. این ارزیابی برای کاربردهایی مانند کنترل پیش‌بین مدل، که مدل را به طول افق کنترل جلو می‌برد تا دنباله بهینه از جابجایی‌های متغیر کنترل‌شده و نقض محدودیت‌های آینده پیدا کند، لازم است و اهمیت پیش‌بینی بدون وابستگی به داده‌های اندازه‌گیری را نشان می‌دهد.

در طول دوره نمونه‌گیری تحقیقات ما، قیمت بیت‌کوین ویژگی‌های یک سری زمانی ناپایدار را نشان داد و توالی تفاوت نتوانست نوع خاصی را شناسایی کند. ما یک چارچوب پیش‌بینی نوآورانه متمرکز بر مدل LSTM را در پاسخ به این چالش‌ها معرفی کردیم. این چارچوب به هدف پیش‌بینی قیمت روزانه بیت‌کوین با ترکیب دو مدل متمایز LSTM هدف داشت.



نتیجه‌گیری و کارهای آینده

بیت‌کوین به عنوان یک منبع تنوع معتبر شناخته شده است که به سرمایه‌گذاران در دوره‌های چالش‌برانگیز پایداری پتانسیل ارائه می‌دهد. ویژگی‌های منحصر به فرد بیت‌کوین باعث می‌شود که امکان ارزیابی دقیق ولاتیلیته آن در دوره‌های کوتاه و طولانی فراهم شود. طراحی پیشنهادی به طور بسیار مؤثری اثرگذار بوده و به پیش‌بینی دقیق قیمت‌های آینده بیت‌کوین کمک کرده و به کاهش ریسک سرمایه‌گذاران کمک می‌کند. اثبات کارایی چارچوب مشخص شده بر روی سری زمانی جایگزین، نشان می‌دهد که این چارچوب پتانسیل بالایی دارد که حتی در شرایط بحرانی نیز بازدهی قابل توجهی داشته باشد. این مطالعه به تحلیل قابل پیش‌بینی بودن رمزارز با بزرگترین سرمایه‌گذاری بازار هدف دارد. روش پیشنهادی می‌تواند به عنوان یک ابزار برای ارزیابی قابلیت پیش‌بینی و کارایی ارزهای دیجیتال مختلف مورد استفاده قرار گیرد و به درک گسترده‌تری از رفتار بازار کمک کند. این تحقیق مسیرهای جدیدی را برای توسعه چارچوب‌های پیش‌بینی جدید با استفاده از اجزا جایگزین مرتبط باز می‌کند.

Ipybn file is attached.
References are attached.