

# **Research article**

# **Prediction of bitcoin stock price using feature subset optimization**

**Shiva Vafadar 810899074**  
**Machine Learning course**  
**Professor Mirzaei**

پیش‌بینی دقیق ارزشی بیت‌کوین برای تصمیم‌گیری‌های سرمایه‌گذاری مطلوب بسیار حیاتی است. 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.





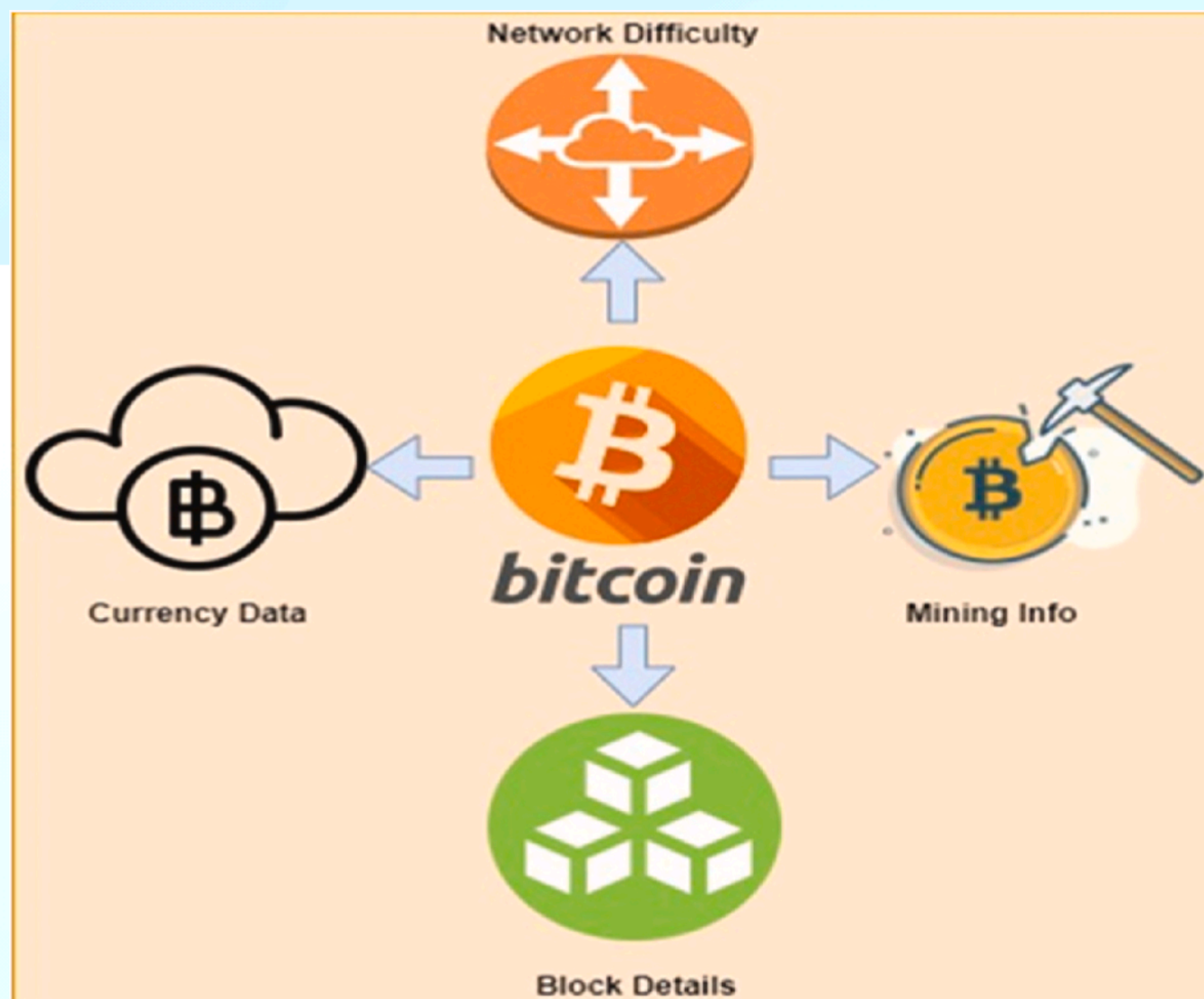
دقیقاً چه چیزی باعث ارزش بیت‌کوین می‌شود؟

چه عواملی بر ارزش بیت‌کوین تأثیر می‌گذارند؟

ارزش بیتکوین نمایانگر اعتقاد سرمایه‌گذاران به ارزشهای رمزنگاری شده در زمینه نوآوری مالی است!

به دلیل نوسانات بیتکوین، تعیین اینکه کدام ویژگی‌ها در زمان پیش‌بینی استفاده شود چالشی است.

این تحقیق به دنبال بدست آوردن بینش‌هایی برای شناسایی که ویژگی‌های داخلی بیتکوین باید برای پیش‌بینی ارزش بیتکوین استفاده شوند.



ویژگیهای بیتکوین:



# روشهای پیشبینی قیمت بیت کوین

- روشهای سنتی

- پیش بینی برحسب LSTM

## روش XGBOOST

XGBoost حاصل پیوند درختهای تصمیم و تقویت گرادیان است و با بهره‌گیری از تکنیکهای پیشرفته، عملکرد درختهای تصمیم را به سطح جدیدی ارتقا داده است. این الگوریتم، با استفاده از توابع هدف و روشهای بهینه‌سازی پیچیده، به‌طور دقیق‌تر و کارآمدتر از درختهای تصمیم ساده پیش‌بینی می‌کند.

### Evolution of Tree Algorithms

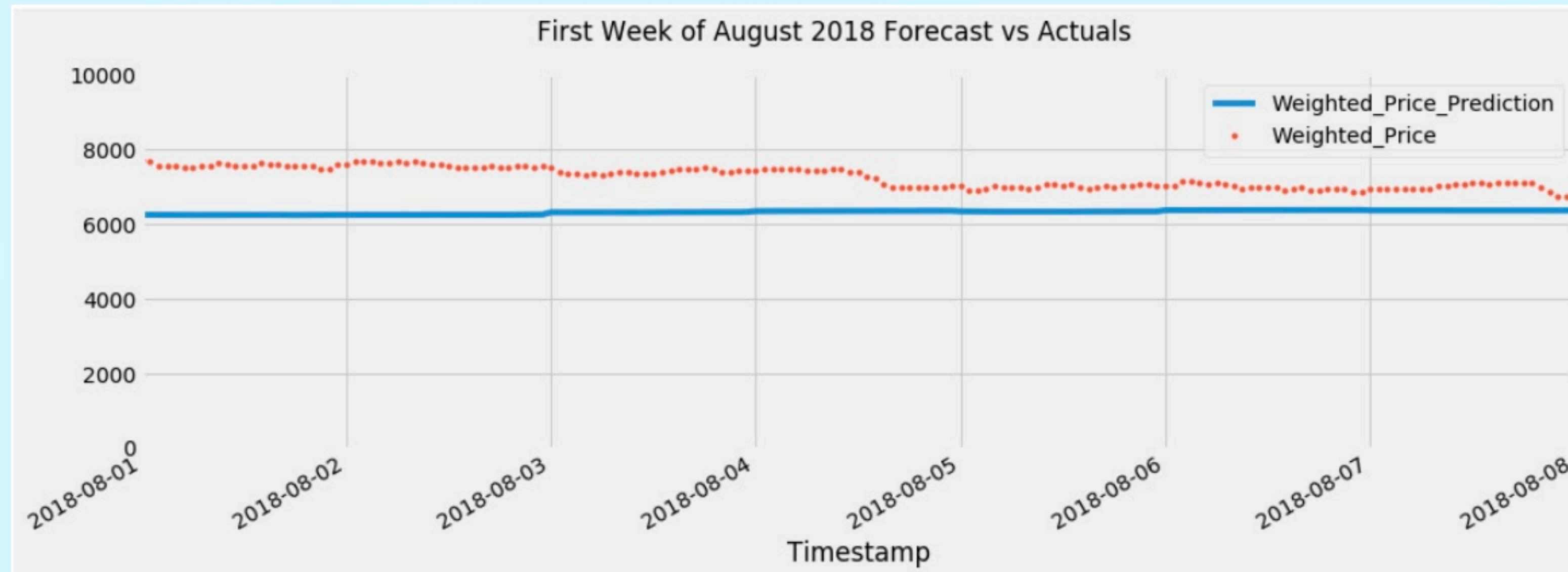




# چرا XGBOOST ؟

XGBoost، به دلیل دقت بالا، کارایی محاسباتی عالی و انعطاف‌پذیری در تنظیم پارامترها برجسته است. این الگوریتم با استفاده از تکنیک‌های پیشرفته مانند پردازش موازی، هرس درختان تصمیم (Tree Pruning) و رگولاریزاسیون (Regularization) به افزایش کارایی و جلوگیری از بیش‌برازش (Overfitting) کمک می‌کند و در نتیجه، در حل مسائل پیچیده و داده‌های ناقص یا نویزدار بسیار مؤثر است.





**MSE=484476.69**

```
model = xgb.XGBRegressor(objective='reg:
linear',min_child_weight=10, booster='gbtree',
colsample_bytree = 0.3, learning_rate
= 0.1,
                        max_depth = 5, alpha = 10,
n_estimators = 100)
model.fit(X_train, y_train,
          eval_set=[(X_train, y_train), (X_t
est, y_test)],
          early_stopping_rounds=50,
          verbose=False) # Change verbose to T
rue if you want to see it train
```

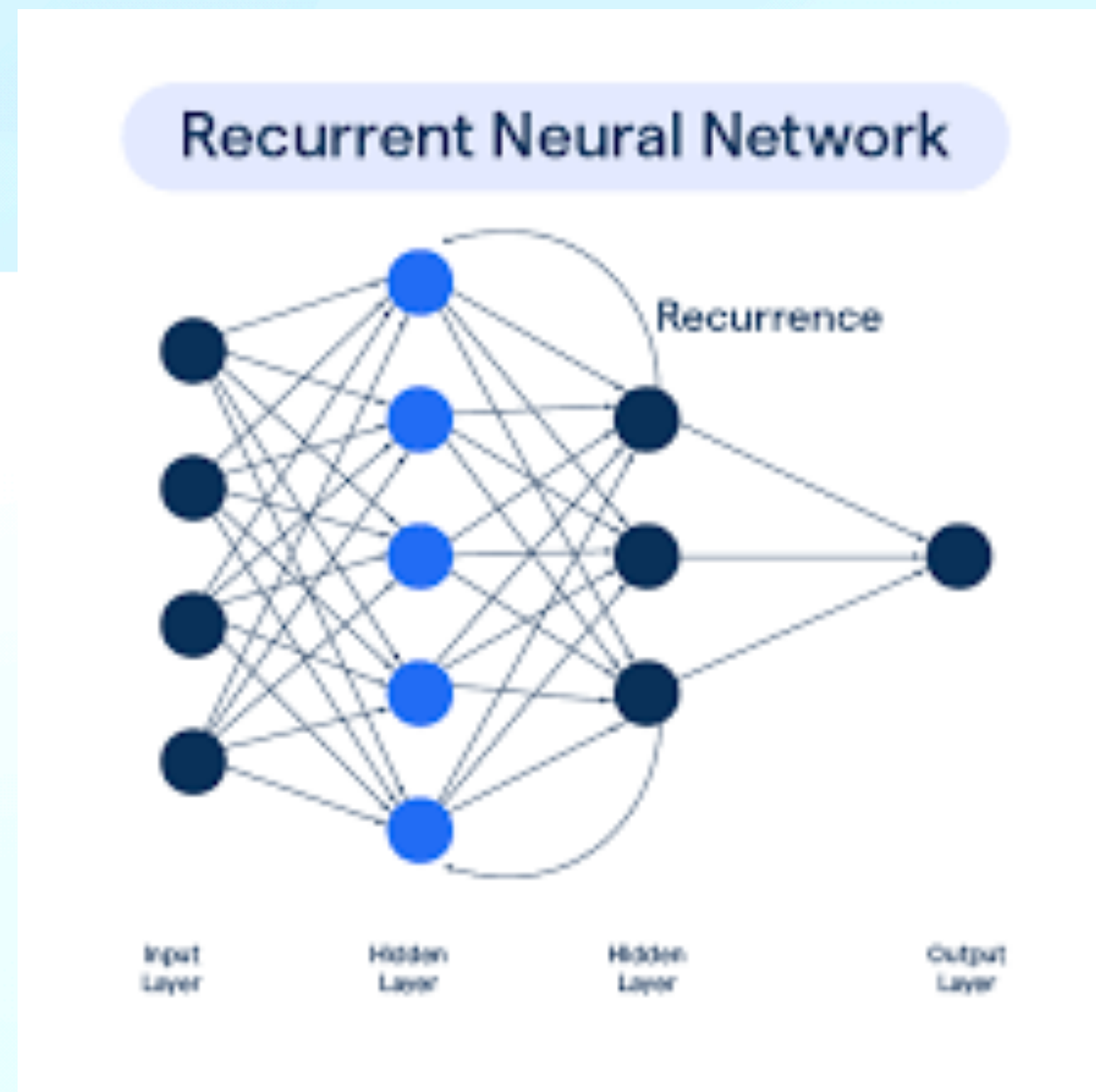
Out[32]:

```
XGBRegressor(alpha=10, base_score=0.5, boo
ster='gbtree', colsample_bylevel=1,
             colsample_bytree=0.3, gamma=0, lear
ning_rate=0.1, max_delta_step=0,
             max_depth=5, min_child_weight=10, m
issing=None, n_estimators=100,
             n_jobs=1, nthread=None, objective='
reg:linear', random_state=0,
             reg_alpha=0, reg_lambda=1, scale_po
s_weight=1, seed=None,
             silent=True, subsample=1)
```



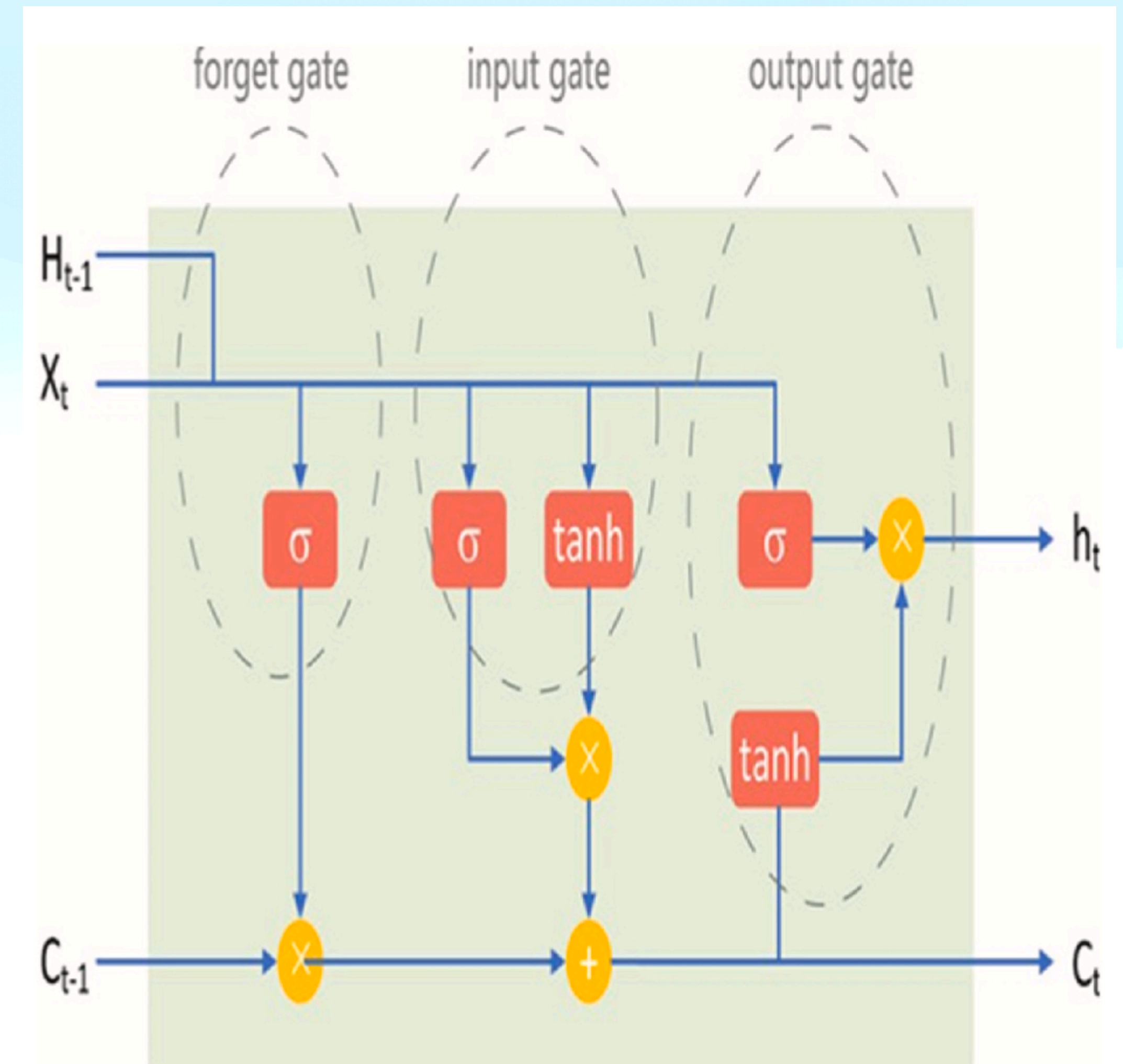
- RNN ها به عنوان شبکه های حفظ و نگهداری اطلاعات عمل می کنند، به ویژه برای کارهایی که شامل توالی هایی مانند تشخیص گفتار و تولید موسیقی است.

- RNN ها با چالش حافظه کوتاه مدت محدود مواجه هستند که به عنوان مشکل گرادیان ناپدید شدن شناخته می شود. این مسئله زمانی به وجود می آید که توالی ها طولانی می شوند و انتقال اطلاعات از مراحل اولیه به مراحل بعدی دشوار می شود.





- شبکه های LSTM، زیرمجموعه ای از RNN ها، به طور خاص برای به خاطر سپردن و پیش بینی داده ها با گرفتن وابستگی های بلندمدت ذاتی در داده های سری زمانی طراحی شده اند.



LSTM با اصلاح پارامترها با استفاده از پنجره های از داده های قبلی به عنوان ورودی، با هدف به حداقل رساندن تفاوت بین مقادیر پیش بینی شده و اندازه گیری شده در مراحل بعدی آموزش داده می شود. الگوریتم های متوالی بر اساس فریم های داده قبلی، یک مقدار بعدی را پیش بینی می کنند.



در زمینه پیش‌بینی قیمت بیت‌کوین، شبکه‌های LSTM از موارد زیر بهره می‌برند تا دقت پیش‌بینی را بهبود بخشند:

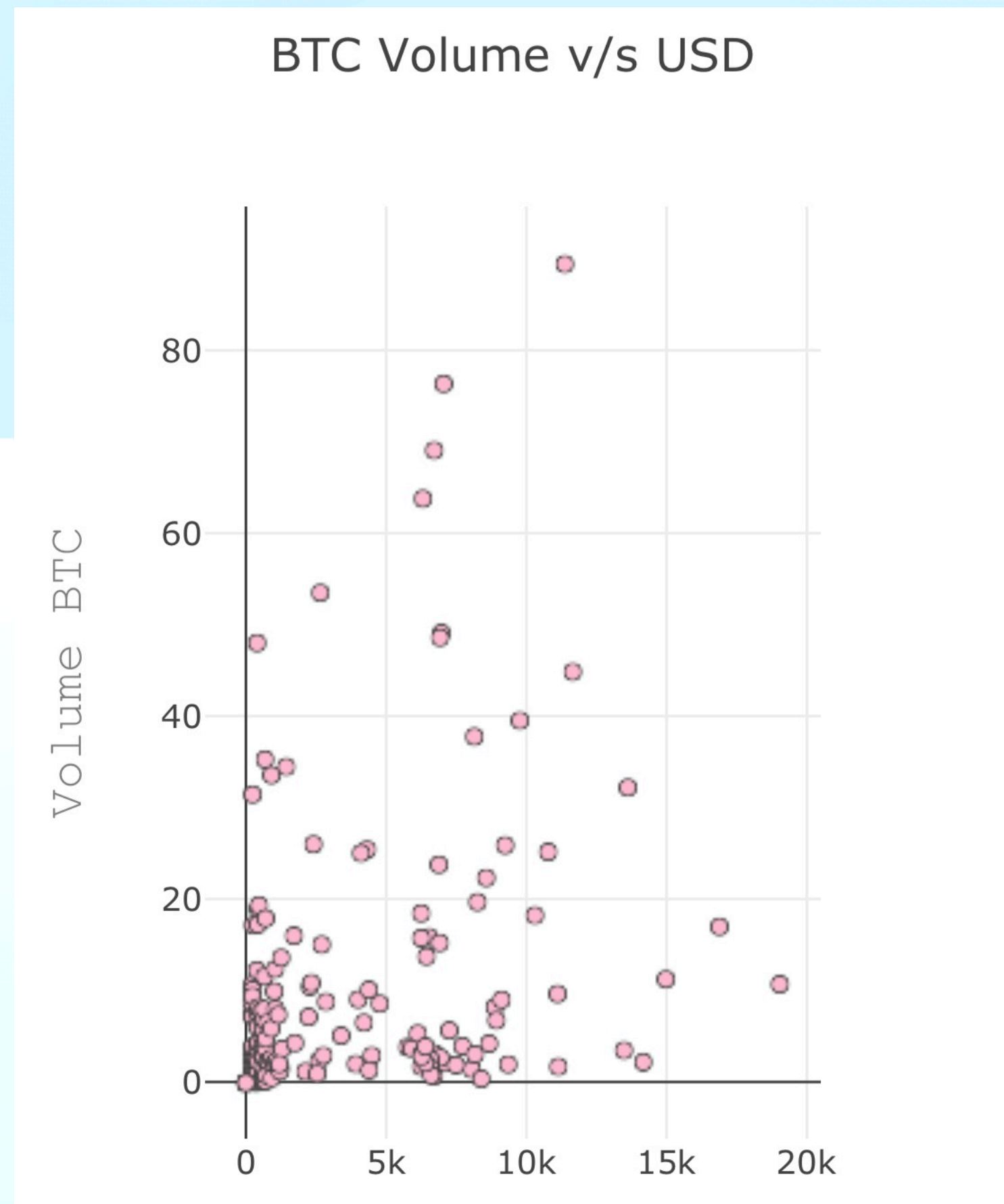
۱. **یادگیری از داده‌های تاریخی:** LSTM‌ها بر روی داده‌های تاریخی قیمت آموزش داده می‌شوند و الگوها و روندهایی که در گذشته رخ داده‌اند را یاد می‌گیرند.

۲. **انتخاب ویژگی‌ها:** با گنجاندن ویژگی‌هایی مانند جزئیات بلوک و حجم معاملات، LSTM می‌تواند نه تنها حرکات قیمت بلکه عوامل اساسی تأثیرگذار بر این حرکات را درک کند.

۳. **کاهش نویز:** از طریق مکانیزم‌های گیت‌بندی، LSTM‌ها می‌توانند نویز را فیلتر کرده و بر روی اطلاعات مهم‌تر تمرکز کنند که برای ایجاد پیش‌بینی‌های قابل اعتماد در بازاری بی‌ثبات مانند ارزهای رمزنگاری شده ضروری است.

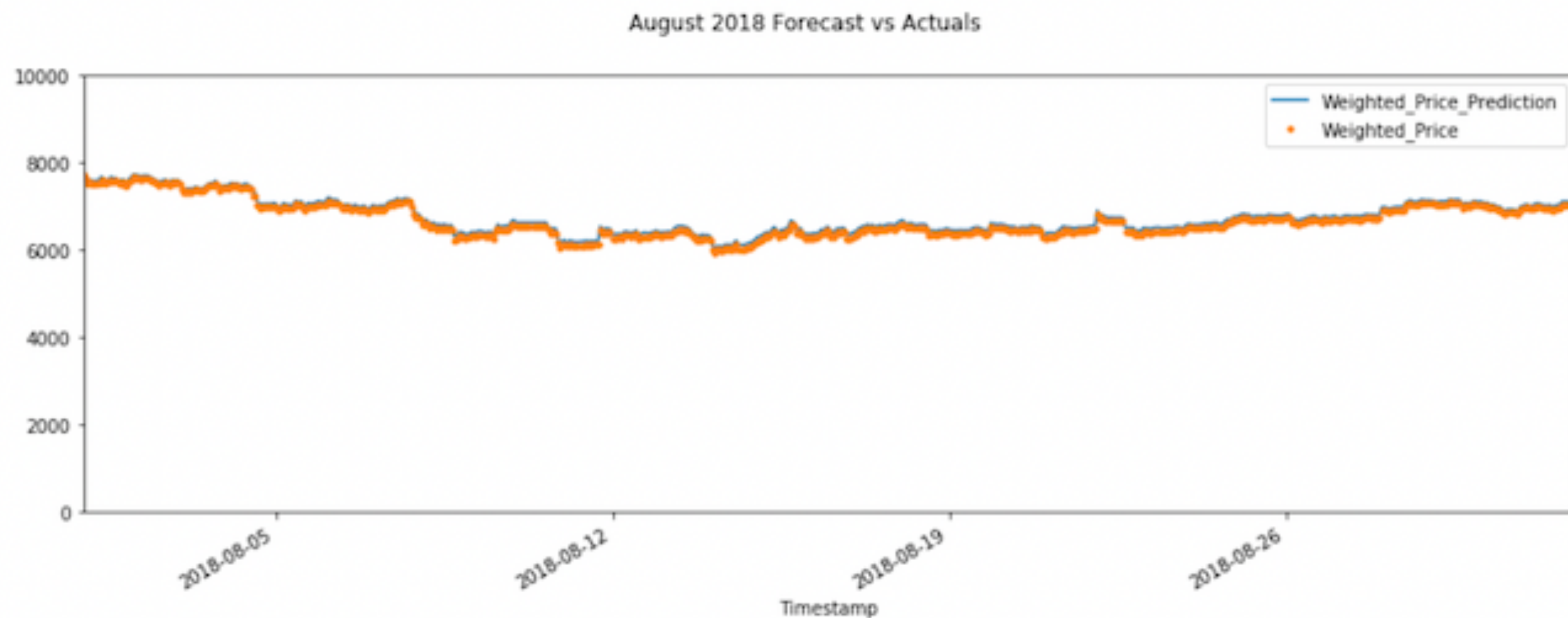
# This Kernel is divided into two parts:-

- \* Data Exploration
- \* Time Series Analysis





# LSTM



```
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)
```

**7439.6052**

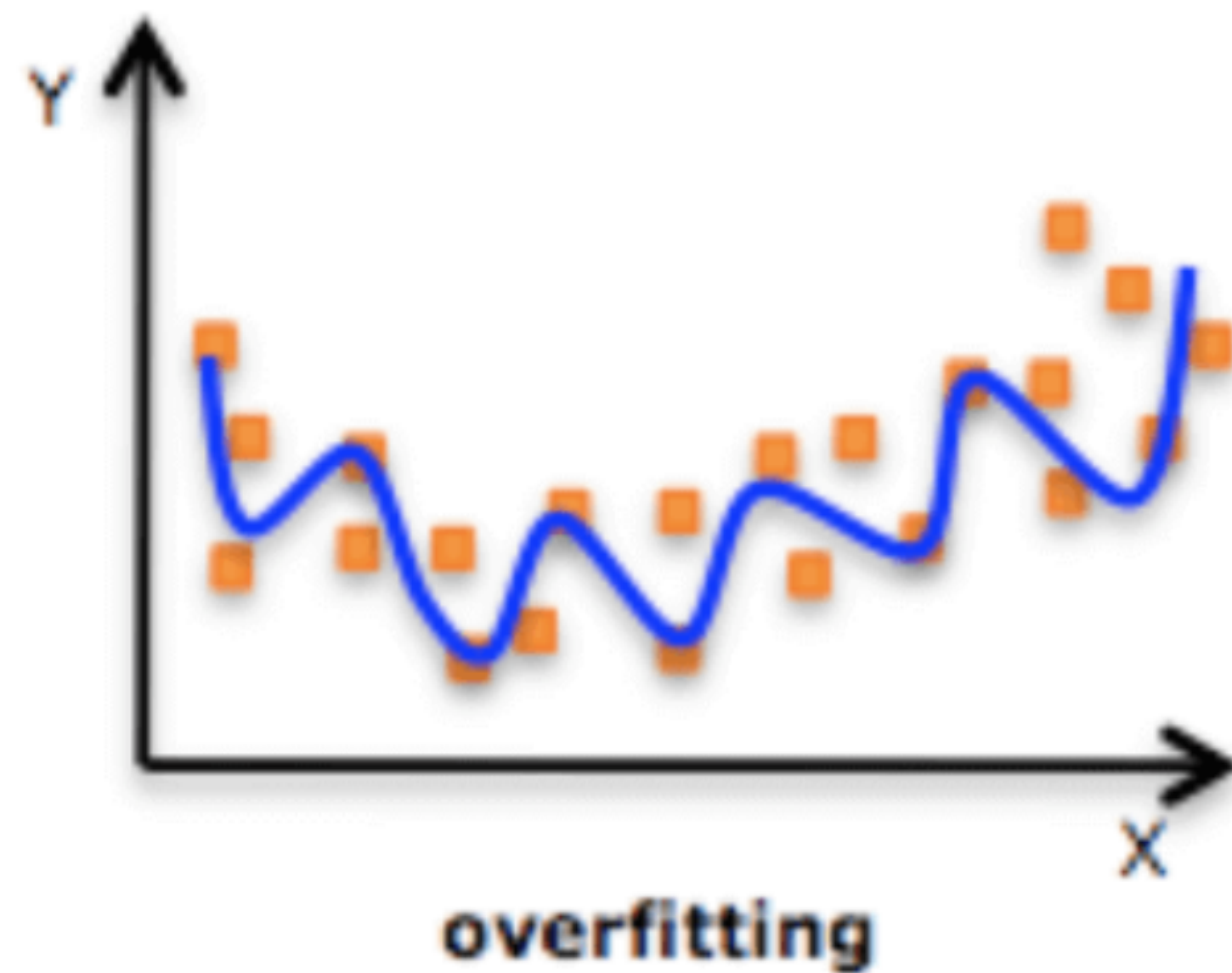
**Mean squared error**



# 85.6052

**Mean absolute error**

# This model is overfitted





می‌توانیم عملکرد مدل‌های بالا را با داده‌های قیمت بیت کوین تاریخی ببینیم. پیش بینی خوب نیست. قیمت بیت کوین بسیار بی ثبات و بسیار تصادفی است و اغلب تحت تأثیر عوامل خارجی (یا اخبار) مانند مقررات ارزهای دیجیتال، سرمایه گذاری ها یا شایعات ساده در رسانه های اجتماعی است. برای اینکه این مدل‌ها عملکرد بهتر و دقیق‌تری داشته باشند، به داده‌های اضافی از اخبار یا رسانه‌های اجتماعی نیاز داریم.





**Thank you for your attention!**