**گزارش پروژه یادگیری ماشین ( پایتون)**

نویسنده: رضا علوی

اردیبهشت 1401

**طرح مساله**

هدف جمع آوری و تحلیل داده های OHLCV برای نماد BTCUSDT از وبسایت صرافی Binance است. در ادامه داده­ها تحلیل میشوند و همبستگی داده ها با داده های روزانه طلا و نفت جهانی بدست آورده میشود. در آخر نیز یک مدل یادگیری ماشین ساده برای پیش بینی قیمت در روز بعد ساخته میشود.

**روند برنامه کامپیوتری دریافت داده ها**

در ابتدا با استفاده از کتابخانه CCXT برای پایتون داده های روزانه دریافت شده و در دیتافریم Pandas ذخیره میشوند. به دلیل محدودیت در ارسال درخواستهای مکرر به صرافی، روشی برای دریافت کامل داده ها از سال2017 تا کنون در برنامه تعبیه شده است. سپس داده ها در دیتابیس MongoDB ذخیره میشوند.

برای بدست آوردن نرخ بازدهی روزانه از فرمول زیر استفاده میکنیم:

برای بدست آوردن همبستگی داده ها، به دو فایل CSV برای طلا و نفت نیاز داریم. در ادامه سه دیتافریم را با هم ادغام کرده و داده های روزهای غیر مشترک را حذف میکنیم. برای بدست آوردن همبستگی از کتابخانه numpy استفاده شده است. همچنین همبستگی بین داده ها با تاثیر تقدم و تاخر هر داده محاسبه شده است.

**مدل یادگیری ماشین**

هدف آموزش یک مدل یادگیری ماشین نظارت شده برای پیش بینی نرخ بازدهی روز بعد با استفاده از نرخ بازدهی پنج روز متوالی است. در ابتدا یاد آوری میکنیم که به احتمال زیاد این مدل در پیش بینی داده های ندیده کارا نخواهد بود. ورودی مدل پنج عدد نرخ بازدهی روزانه و خروجی مدل نوع کلاستر بازدهی مثبت، نزدیک صفر و منفی است. برای دسته بندی مقادیر خروجی از انحراف از میانگین کل نرخ های بازدهی با احتساب ضریبی از انحراف معیار استفاده شده است.

با توجه به توالی داده ها، استفاده از مدلهای شبکه عصبی بازگشتی و کتابخانهTensorflow بهترین گزینه است. اما با توجه به سادگی متغیر ها از ساختار شبکه عصبی MLP و کتابخانه مقدماتی یادگیری ماشینSKLearn استفاده شده است. با توجه به مقادیر ورودی ها، پیش پردازش داده ها صورت گرفته و توجه داریم که لایه های شبکه عصبی از تابع راه اندازی RELU استفاده کنند. پس از تقسیم داده ها به دو قسمت Train و Test مدل را آموزش میدهیم. سپس مدل را تست میکنیم و پیشبینی ها را تحلیل میکنیم. خروجی به دست آمده در صورتی مطمئن تلقی میشود که انحراف کافی از میانگین پیشبینی ها داشته باشد. نمونه ای از خروجی برنامه به همراه confusion matrix[[1]](#footnote-1) را مشاهده میکنید:

From 514 predictions: 18 positive and 35 Negetive

With dev=1.5 confidence level, 10 % of predictions are confident enough

Results on test data: 14TP, 4FP, 22TN, 13FN

Accuracy = 67.9%

This Strategy returns 56% Profit.

**نتیجه گیری**

با وجود دقت بدست آمده 67 درصدی روی داده های تست، نمیتوان از مدل به عنوان یک پیشبینی کننده قیمت خوب انتظار داشت. باید توجه کنیم که نرخ بازدهی از سال 2017 در اکثر روزها مثبت بوده است و زمانی دقت مناسب خواهد بود که سیستم نرخ بازدهی بالاتر از بازدهی 2017تا کنون را داشته باشد.

1. <https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826> [↑](#footnote-ref-1)