

به نام خدا

توضیحات کدهای ارسال شده  
پیش‌بینی قیمت نفت

کارآموزی سازمان برنامه و بودجه - تابستان ۱۴۰۱

طیبه محبی

## فهرست

### Contents

۳	..... LSTM2-useful
۳	..... DNN first try
۳	..... DNN!
۴	..... linear regression
۴	..... RandomForest
۴	..... SVM1
۴	..... XGBoost
۴	..... A long and meaningless code
۵	..... A long and meaningless code2
۵	..... Prophet
۵	..... VAR
۵	..... LR, ANN
۶	..... BPNN
۶	..... BPNN revised
۶	..... Backprop first try
۶	..... Tsfresh
۷	..... Semi-final BPNN

کدها همگی قابل ران کردن می باشند (مگر چند فایل کد که در اول کد این مسئله بیان شده است).

## LSTM2-useful

یکی از مدل های اولیه پیشبینی. هدف پیشبینی ماهانه است، و میبینیم که نتیجه اصلا مناسب نیست. از این کد استفاده خاصی نمی توان کرد، اما نمودارهایی که می کشد میتوانند مفید باشند.

## DNN first try

اولین تلاش برای اجرای شبکه عصبی روی داده ها! تنها فیچر دیتاست، قیمت ماهانه wti است. دیتاها با لگ یا پنجره ۲ تقسیم می شوند و سائز تست، ۳۲ ماه است. در این کد، پنجره های ساخته شده را شافل میکنیم تا نوسان مدل کاهش یابد. نتیجه قابل قبول است و مدل با خطای حدود ۴ دلار پیشبینی را انجام می دهد. اما این مدل فقط برای پیشبینی ماه به ماه اعتبار دارد.

## DNN!

اولین تلاش برای پیشبینی مستقیم هشت ماه بعد توسط شبکه عصبی. کد قابل ران کردن می باشد و فیچرهای استفاده شده، قیمت ماهانه WTI، GDP آمریکا و DJIA (شاخص بورس ۳۰ شرکت بزرگ آمریکا) می باشد.

متأسفانه نتیجه اصلا قابل قبول نیست. کد قابل ران کردن است اما میبینیم نتیجه غیر منطقی است. ممکن است این مسئله به دلیل اسکیل و نرمال نکردن دیتاها باشد. اما نتیجه در هر صورت به اندازه فایل قبلی شبکه عصبی قابل قبول نیست.

### نکات قابل تغییر:

در دیتاست داده شده، فقط ستون قیمت را نگه داریم و دو ستون دیگر را حذف کنیم. همچنین در سلول سوم، نام ستون هایی که حذف شده باید حذف شوند.

میبینیم نتیجه تا حدود  $rmse = 23$  بهتر می شود.

اما باز هم نتیجه، میانگین شش ماهانه نیست.

پیشنهاد: ۶ پیشبینی انجام دهیم، هر کدام برای ماه بعد، دو ماه بعد و الی آخر. سپس میانگین بگیریم.

## linear regression

یکی از کد های اصلی. ابتدا بهترین دیتاست انتخاب شده و سپس پیشبینی انجام میشود. کد دارای کامنت می باشد. به دیتاست توجه کنید، ستون **real** درواقع میانگین شش ماه بعد است، یعنی خروجی این مدل پیشبینی شش ماهه است. این مدل در حالت اولیه، تک متغیره است. اما به سادگی میتوان به فایل اکسل ستون های جدید اضافه کرد و کافیسیت به کامنت های کد توجه کنید تا به راحتی مدل به **MLR()** تبدیل شود.

## RandomForest

توضیحات دقیقا مثل رگرسیون خطی می باشند. الگوریتم، رندوم فارست است این کد هم قابلیت چند متغیره شدن دارد.

## SVM1

توضیحات مثل دو مورد بالا. الگوریتم، SVM است.

## XGBoost

مانند موارد بالا. الگوریتم، xgboost است.

## A long and meaningless code

این کد کاملا تمرینی بوده و هیچ کاربردی ندارد. تلاش در نوشتن یک LSTM کاربردی و یک شبکه عصبی از پایه بوده است.

## A long and meaningless code2

این کد هم کاملاً تمرینیست و تلاش برای نوشتن یک شبکه عصبی از پایه بوده.

### Prophet

این کد برای استفاده دستی از پکیج prophet فیس بوک نوشته شده است. به چند نکته توجه کنید:

نمودارهای پیشبینی در هر مرحله با دیتاست های مختلف کشیده شده اند و نتایج چاپ شده اند. دقت کنید در این کد هم دیتاست متغیر داریم.

داده های تست سه عدد هستند و باید لیبل داده های تست را دستی به مدل بدهیم. در بلاک سوم کد این موارد رعایت شده اند، تغییر آن و متناسب کردنش با پیشبینی های دیگر کار ساده ای است. پیشبینی داده شده، برای شش ماهه منتهی به می، ژوئن و ژولای ۲۰۲۲ می باشد.

### VAR

این کد، مجموعه ای از تست های آماری را روی دیتای داده شده انجام میدهد. کد نسبتاً طولانیست اما نتایج کاملاً واضح هستند و با یک بار ران، همه چیز مشخص می شود. در مرحله آخر تلاش بر این بود که یک پیشبینی پیشرفته آماری نیز داشته باشیم، اما شکست خوردم. پس از این کد فقط برای انجام تست های آماری میتوان استفاده کرد

### LR, ANN

این کد را برای مقاله ای نوشتم که در پاورپوینت لیتزرچر ریویو، شماره اول است. تمامی موارد عیناً طبق مقاله بوده و فقط دیتاست تفاوت دارد. کد قابل ران کردن است. دقت کنید با ران های مختلف، نتایج متفاوتند که این طبیعی

است. بهترین مدل اینجا سیو نشده، اما بهترین  $rmse$  به دست آمده 8.192 است. دقت کنید این مدل پیشبینی ماهانه انجام میدهد، نه شش ماهه.

## BPNN

مدل تمرینی و بی فایده. میبینیم پیشبینی اصلا خوب نیست و  $rmse$  حدود ۴۰ داریم.

## BPNN revised

این کد هم تمرینی بوده و ارزشی ندارد.

## Backprop first try

تمرینی و بدون ارزش.

## Tsfresh

این مدل، یکی از مدل های اصلی پیشنهادی بنده است. ابتدا با پکیج Tsfresh فیچرها استخراج می شوند و سپس با یک شبکه عصبی، پیش بینی انجام می شود. دقت پیشبینی عالی نیست،  $rmse$  حدود ۳۵ نتیجه نهایی است. اما واقعیت این است که برای پیشبینی شش ماهه، بخصوص در شش ماهه های اخیر که نوسان بالا بوده، نتیجه بدی نیست و فیچرهای استخراج شده، موارد جالبی دارند. یکی از ویژگی های جالب مدل، این است که نوسان هایی که پیشبینی شبکه عصبی خالص دارد را به شدت خنثی می کند.

در کل این مدل جای کار بیشتری دارد و امید بالایی به آن می رود. دقت کنید در صورت تغییر دیتاست، پارامترها باید متناسب شوند. ولی فرم کلی کد به همین شکل است.

دقت کنید که در این کد، دیتاست متغیر انتخاب نکرده ایم و در صورت هوشمندی در انتخاب دیتاست، نتیجه جای بهبود بسیار دارد.

## Semi-final BPNN

این هم یکی از کدهای نهایی می باشد. من برای تست، دیتاست تغییر یافته ای که در موارد رگرسیون خطی نتیجه بهتری میداد را به عنوان ورودی مدل دادم. واضح است که چون مدل دیپ است، تفسیر نتایج به سادگی تفسیر رگرسیون خطی نیست.

مسئله اصلی ای که اینجا وجود دارد، نوسان مدل است. با چند بار ران کردن مدل، از  $10\text{rmse}$  داریم تا  $30$ . البته در کل مدل دیپ، نیاز به دیتای بیشتری دارد.

برای استفاده از کد، دقت کنید که در دیتاست، ستونی که میخواهیم پیشبینی کنیم باید اولین ستون باشد. همچنین در بعضی جاهای کد نیاز به تغییرات کوچکی داریم (در صورت تغییر دیتاست) که در این مورد به کامنت ها توجه کنید.

این مدل به خودی خود به دلیل نوسان زیاد، قابل استفاده نیست. اما میتوان بهترین مدل را ذخیره کرد (طبق کامنت ها) و از همان استفاده کرد.