

### «مسئله پیش‌بینی قیمت نفت»

در این پروژه، شما باید با استفاده از تکنیک‌هایی که تا بحال یاد گرفته‌اید، تلاش کنید که قیمت نفت را برای روزهای آینده پیش‌بینی نمایید. این پروژه برگرفته از اطلاعات واقعی قیمت نفت برنت است و پیش‌بینی دقیق قیمت می‌تواند کمک شایانی به نحوه قیمت‌گذاری نفت ایران نماید.

**بخش ۱)** دیتاست `brent_oil_weekly_30_years.csv` را لود کنید. هر ردیف در این دیتاست حاوی اطلاعات قیمت‌های معامله‌شده در یک روز خاص می‌باشد. ستون‌های این دیتاست عبارتند از:

- `Date`: تاریخ مربوط به اطلاعات یک ردیف
- `Price`: میانگین قیمت معامله‌شده در یک روز
- `Open`: قیمت شروع در یک روز
- `High`: بالاترین قیمت معامله‌شده در یک روز
- `Low`: پایین‌ترین قیمت معامله‌شده در یک روز
- `Vol.`: حجم معاملات انجام‌شده در روز
- `Change %`: درصد تغییر قیمت نسبت به تاریخ قبلی (هفته پیشین)

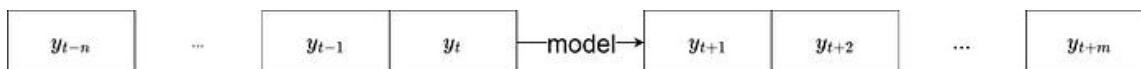
پس از لود کردن دیتاست، تایپ هر ستون را به‌طور مناسب تغییر دهید. بطور مثال، ستون `Date` حتما باید از جنس تاریخ باشد. همچنین، ردیف‌ها و ستون‌های بدون مقدار را نیز بطور مناسب رسیدگی نمایید (`drop` کردن در این مسئله لزوماً راه‌حل منطقی نیست). سپس به ازای هر ستون، نمودار سری‌زمانی آن را رسم نمایید.

**بخش ۲)** برای پیش‌بینی یک سری‌زمانی، دو روش متداول وجود دارد:

- پیش‌بینی سری زمانی در زمان  $t+1$  با استفاده از  $n$  نقطه قبل از آن همانند شکل زیر:



- پیش‌بینی سری‌زمانی در زمان‌های  $t+1$  تا  $t+m$  با استفاده از  $n$  نقطه قبل از آن‌ها همانند شکل زیر:



در [این لینک](#) تصویرسازی خوبی برای این کار انجام شده است.

در این بخش شما باید تابعی بنویسید که با دریافت دیتافریم اولیه و پارامترهای  $m$  و  $n$ ، یک دیتافریم جدید را بعنوان خروجی برگرداند که به ازای یک تاریخ  $t$ ، داده‌های  $n$  هفته قبل و  $m$  هفته بعد در هر ردیف قرار گرفته باشد. بطور مثال به ازای  $m=2$  و  $n=2$ ، دیتافریم جدید باید شامل ستون‌های زیر باشد (نامگذاری ستون‌ها می‌تواند به‌نحو دیگری باشد):

- Date: تاریخ مربوط به یک سطر (t)
- Vol.\_t-1, Low\_t-2, Low\_t-2, High\_t-2, High\_t-1, Open\_t-2, Open\_t-1, Price\_t-2, Price\_t-1
- Vol.\_t-2, Change %\_1, Change %\_2: اطلاعات مربوط به دو هفته پیشین نسبت به زمان t
- Price\_t+1, Price\_t+2: اطلاعات قیمت مربوط به دو هفته پس از زمان t

دقت کنید که به ازای یک t خاص و برای روزهای پس از زمان t، صرفاً نگهداری مقادیر Price کافیست.

**بخش ۳)** برای ارزیابی مدل‌ها، اطلاعات قیمت‌گذاری نفت را از سال ۲۰۱۹ بعنوان داده ارزیابی و داده‌های قبل از ۲۰۱۹ را بعنوان داده آموزش در نظر بگیرید.

**بخش ۴)** با استفاده از تابعی که در بخش ۲ توسعه داده‌اید، یک دیتاست بسازید. در این بخش، مقدار m را 1 و مقدار n را 8 انتخاب نمایید. سپس سعی کنید که فقط با استفاده از n مقدار Price متوالی پیشین، مقدار Price را برای هفته آینده پیش‌بینی نمایید. برای پیش‌بینی قیمت می‌توانید از انواع مدل‌های مختلف نظیر LinearRegression, DecisionTreeRegressor, SVR, GradientBoostingRegressor و ... استفاده نمایید. با محاسبه متریک‌های MAE و MSE، مدل‌های خود را ارزیابی نمایید.

**بخش ۵)** پیش‌بینی‌های انجام‌گرفته توسط هرکدام از مدل‌ها را روی یک نمودار به همراه مقادیر واقعی رسم نمایید.

**بخش ۶)** با تغییر مقدار n و ساختن مدل‌هایی یکسان، به ازای کدام مقدار n، عملکرد مدل بیشینه می‌شود؟

**بخش ۷)** با در نظر گرفتن ویژگی‌های دیگر دیتاست نظیر Low، High و ...، مقدار Price را برای هفته بعد پیش‌بینی نمایید. آیا اضافه کردن ویژگی‌های جدید عملکرد بهتری را در پیش‌بینی قیمت رقم زد؟

**بخش ۸)** با استفاده از روش Forward Selection، حداکثر ۱۲ ویژگی را بیابید که انتخاب آن‌ها در آموزش مدل به بهترین نتیجه منتهی بشود. اختلاف عملکرد این مدل با مدل بخش ۷ چقدر است؟

**بخش ۹)** مدلی بسازید که با دریافت اطلاعات مربوط به ۱۲ هفته متوالی، قیمت را برای ۴ هفته آتی پیش‌بینی نماید. بهترین مدل برای این کار چه ویژگی‌هایی دارد؟ با استفاده از Grid Search پاسخ دهید. آیا عملکرد این مدل از مدل‌هایی که برای پیش‌بینی یک تنها هفته آتی آموزش دیده‌اند، بهتر است؟

**بخش ۱۰)** شما در کل کدام یک از مدل‌هایی را که تا بحال استفاده کرده‌اید، برای حل این مسئله پیشنهاد می‌دهید؟

**بخش ۱۱ و امتیازی)** با استفاده از تکنیک Empirical Mode Decomposition، نوسانات قیمت (Price) را به نوسانات بلندمدت، میان مدت و کوتاه مدت تجزیه کنید و این نوسانات را بعنوان ویژگی در دیتافریم اضافه کنید. آیا اضافه کردن این ویژگی‌ها می‌تواند به بهبود پیش‌بینی مدل منجر بشود؟ برای استفاده از EMD از پکیج EMD-signal استفاده نمایید.

**بخش ۱۲ و امتیازی)** چند مدل مختلف رگرسیونی را برای حل مسئله پیش‌بینی قیمت نفت استفاده نموده و با استفاده از تکنیک Model Stacking، مدل پیش‌بینی بهتری ارائه دهید.

«موفق باشید»