* دریافت فایل داده ها (data set)

یک فایل csv که دارای ۱۵ ستون میباشد که به ترتیب از چپ به راست اطلاعات زیر را شامل میشوند:

* تاریخ
* نام نماد
* نوع امنیت
* قیمت لحظه بسته شدن روز قبل
* قیمت بازگشایی روز
* بیشترین قیمت روز
* کمترین قیمت روز
* قیمت آخرین معامله
* قیمت لحظه بسته شدن
* حجم وزنی میانگین قیمت
* حجم
* حجم معاملات
* معاملات ( خالی میباشد)
* حجم تحویل
* درصد تحویل
* ذخیره کردن اطلاعات هر ستون در یک لیست و نمایش نمودارهای قیمت بازگشایی ، قیمت آخرین معامله و قیمت هنگام بسته شدن ( open, last, close) برای شفاف سازی (visualization) داده ها.
* نرمال سازی داده ها و ساخت یک لیست برای نگه داری feature ها و label ها به نام features که در هر ایندکس آن به ترتیب قیمت بسته شدن روز قبل ، بالا ترین قیمت ، پایین ترین قیمت روز ، قیمت آخرین معامله ، قیمت لحظه بسته شدن میباشد. ( با توجه به نتیجه پیش بینی عدم استفاده از بالاترین و پایین ترین قیمت باعث نتیجه بهتر میشود.)
* Text

  Description automatically generatedتقسیم کردن داده ها به دو دسته train و test که 80 درصد داده ها برای train و باقی برای test استفاده میشوند.

متد train\_test\_split عمل جداسازی داده ها را انجام میدهد که نتیجه را در آرایه های متناظر ذخیره میکند ، که در این متد به طور پیش فرض داده ها را پخش و جا به جا (shuffle) میکند.

* استفاده از الگوریتم KNN (k-nearest neighbors) برای پیش بینی که به ازای مقدار k از ۱ تا ۳۲ مقادیر مختلف در یک دیکشنری ذخیره شدند.

Text

Description automatically generatedبه ازای هر k مقدار accuracy توسط متد accuracy\_score بدست آمده و چاپ شده اند.