* 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:

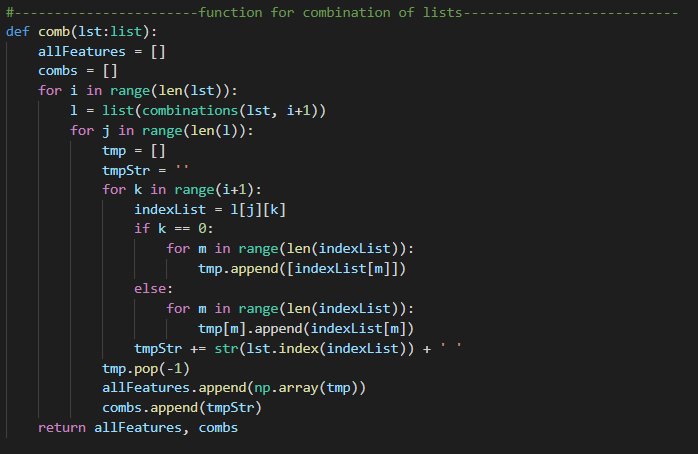
* تاریخ
* نام نماد
* نوع امنیت
* قیمت لحظه بسته شدن روز قبل
* قیمت بازگشایی روز
* بیشترین قیمت روز
* کمترین قیمت روز
* قیمت آخرین معامله
* قیمت لحظه بسته شدن
* حجم وزنی میانگین قیمت
* حجم
* حجم معاملات
* معاملات ( خالی میباشد)
* حجم تحویل
* درصد تحویل
* Text

  Description automatically generatedذخیره 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.

برای آکاهی و فهم نتوع داده ها از شفاف سازی استفاده میشوند که باعث ایجاد دید بهتر به داده ها میشود.

* نرمال سازی داده ها برای تبدیل محدوده ی همه داده ها به مقداری بین 0 و 1 که در این صورت داده های بسیار بزرگ یا بسیار کوچک نسبت به بقیه داده ها ، باعث عملکرد غلط برنامه نمیشود و همه داده ها در یک محدوده قرار میگیرند. فرمول مورد استفاده برای نرمال سازی به صورت زیر میباشد:

Xnormal = (X - Xmin) / (Xmax - Xmin)

* در متد com که به عنوان ورودی یک لیست میگیرد که شامل ویژگی های بالاترین قیمت، پایین ترین قیمت، آخرین معامله، قیمت بسته شدن، نسبت حجم به کل معاملات و نسبت قیمت بسته شدن به آخرین معامله دخیره شده اند و داخل این متد همه ترکیب های ممکن از feature ها را تشکیل میدهیم و به عنوان یک لیست بر میگرداند و خروجی دوم آن یک رشته میباشد که نشان دهنده ترتیب feature های موجود در ایندکس متناظر میباشد.
* متد createModel که برای تبدیل داده ها به دو دسته train و test که روی داده های train به کامپیوتر آموزش میدهیم و روی داده های test چک میکنیم که آموزش ما چه نتیجه ای داشته و برنامه ما چه مقدار از داده ها را توانسته درست تشخیص دهد. برای این آموزش از الگوریتم KNN استفاده میکنیم که این الگوریتم به این صورت عمل میکند که k داده نزدیک داده اضافه شده را بررسی میکند و تعداد هرکدام بیشتر بود به آن گروه اضافه میشود( در این پروژه مقدار Positive یا Negative ) میگیرد. در این متد پس از اجرای الگوریتم KNN و پیش بینی، آن را در یک dictionary ذخیره میکند و چاپ میکند و در آخر بیشترین درصد پیش بینی و ترتیب featureهای انتخابی را چاپ میکند و نمودار آنرا رسم میکند که در نمودار نقاط آبی پیش بینی درست هستند و نقاط قرمز نقاطی هستند ک به اشتباه پیش بینی شده اند.

