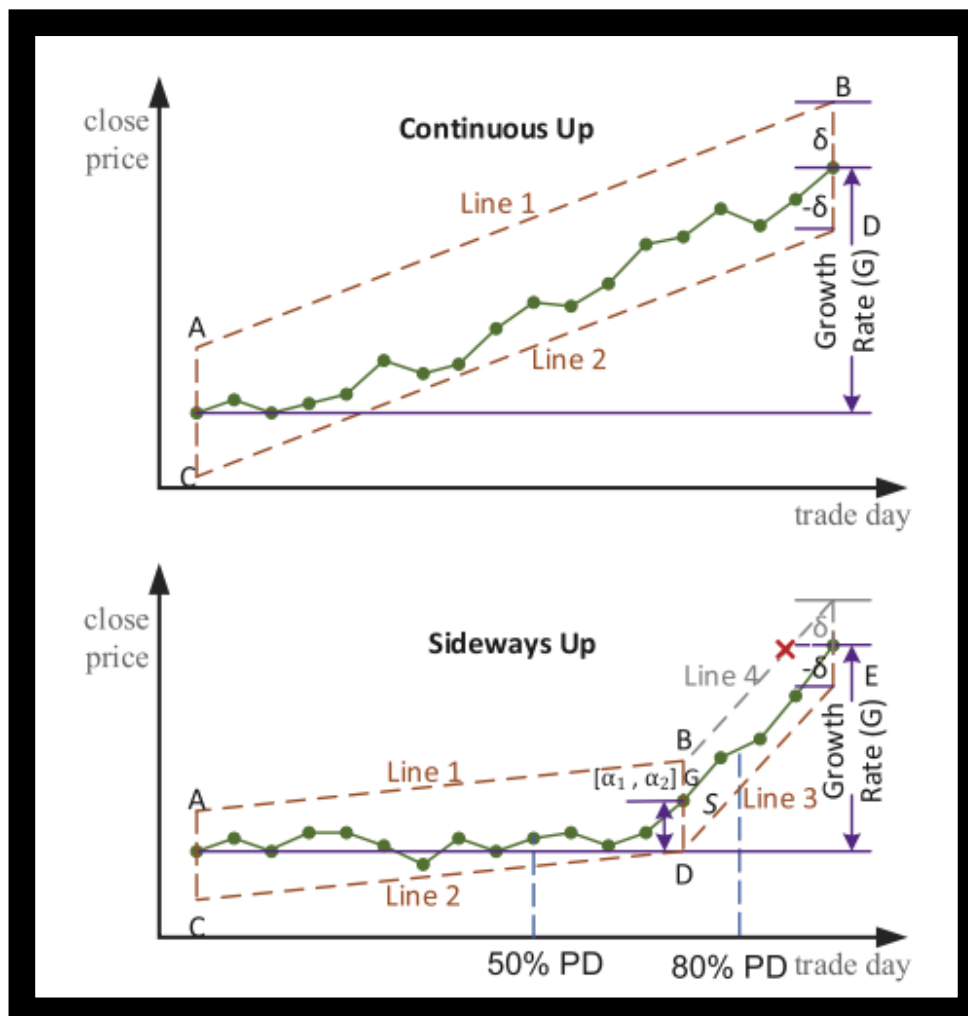
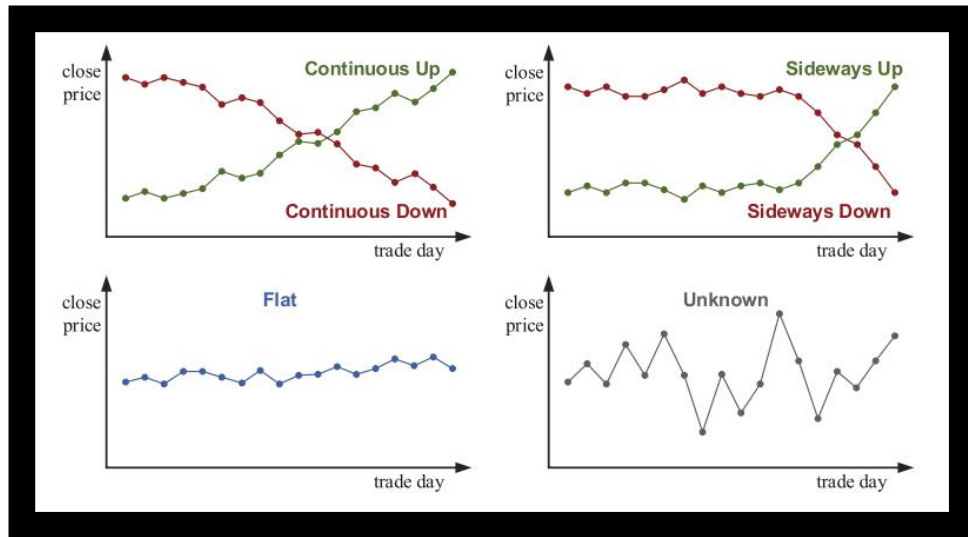


- ❖ روند کلی: این الگوریتم برای پیش بینی روند رشد بازار بورس (چین) میباشد که داده ها را به بازه های ۱۸ روزه تقسیم می کنیم و سپس پیش بینی می کنیم که بازه بعدی که به ما داده می شود جزو کدام کلاس می باشد.
- ❖ در مرحله اول اطلاعات مربوط به سهام را از فایل ADANIPTS خوانده و ستونهای close, open, high, low, volume را در لیست هایی به همان اسم ذخیره شده اند.
- ❖ در مرحله دوم همه feature ها را در یک لیست ذخیره کرده که در ایندکس i ام آن اطلاعات آن روز در آن میباشد.
- ❖ در مرحله سوم داده ها را به گروه هایی تقسیم کرده که این تقسیم بندی به این صورت می باشد که همه دوره های ۱۸ روزه ممکن (۱۸ روز متوالی) یک گروه در نظر گرفته شده که هر یک از این گروه ها یک کلیپ (clip) نامیده می شوند.
- ❖ در مرحله چهارم به هر کلیپ یک کلاس اختصاص داده ام که این کلاس ها برابر Continuous UP/Down, Sideway Up/Down, Unknown, Flat می باشند. که در ادامه هر کدام از آنها توضیح داده شده است.
- ✓ **Continuous Up/Down**: ابتدا نرخ رشد را محاسبه می کنیم سپس چهار نقطه به اندازه ۰.۲ نرخ رشد در بالا و پایین نقطه ابتدای کلیپ و انتهای کلیپ مشخص می کنیم و یک مستطیل تشکیل می دهیم. سپس چک می کنیم اگر تمام نقاط کلیپ داخل مستطیل قرار می گرفتند آنگاه اگر نقطه انتها کلیپ بالاتر از نقطه ابتدا کلیپ بود Continuous Up و اگر پایین تر بود Continuous Down میباشد.
- ✓ **Sideway Up**: ابتدا نرخ رشد را محاسبه می کنیم سپس از نقطه اول شروع می کنیم و به جلو میرویم و تا آخرین جایی که با رسم مستطیل (که در بالا گفته شد) تمام نقاط داخل مستطیل قرار می گیرند جلو میرویم و مستطیل را رسم می کنیم و سپس به اندازه ۰.۲ پایین تر از نقطه آخر کلیپ یک نقطه دیگر رسم می کنیم و به نقطه ای که در مرحله قبل پیدا کردیم وصل می کنیم و اگر تمام نقاط بالای این خط بودند آنگاه این کلیپ Sideway Up می باشد.
- ✓ **Sideway Down**: مانند بالا عمل می کنیم اما با این تفاوت که نقطه پنجم بالای نقطه آخر کلیپ می باشد و به نقطه ای که برای مستطیل پیدا کردیم وصل می کنیم که اگر تمام نقاط پایین تر از این خط بودند آنگاه این کلیپ Sideway Down می باشد.
- ✓ **Unknown**: اگر کلیپ Continuous, sideway Up, sideway Down نباشد آنگاه Unknown می باشد.
- ✓ **Flat**: در صورتی که هیچ کدام از حالات بالا نباشد آنگاه Flat می باشد.
- ✓ توابع IsSidewayUp و IsSidewayDown چک می کند که کلیپ از نوع side way Up/Down باشد یا خیر.



❖ در مرحله پنجم احتمال رخداد (P_m) هر کدام از کلاس های Up, Down, Unknown, Flat در بین داده های ما چقدر است را محاسبه و در لیست P_m ذخیره شده.

```
#-----find probability to occurrence each class-----
P_m = []
P_m.append(len(class0) / len(classss))
P_m.append(len(class3) / len(classss))
P_m.append(len(class1) / len(classss))
P_m.append(len(class2) / len(classss))
```

❖ در مرحله ششم مقدار Entropy را محاسبه می کنیم که مقدار آن برابر جمع مقدار $P_m \log_2 P_m$ برای هر کدام از کلاس ها می باشد.

```
#-----calculate Entropy-----
tmp = 0
for i in range(len(P_m)):
    if P_m[i] != 0:
        tmp += P_m[i]*math.log2(P_m[i])
Entropy = -1*tmp
```

❖ مرحله هفتم محاسبه مقدار D_i می باشد که برابر مقدار فاصله هر عنصر (روز) با عناصر دیگر هست که توسط متد $\text{dist}()$ محاسبه می شود و سپس به صورت نزولی مرتب (sort) می شوند و متناسب با همان تغییر در این لیست ترتیب هر ایندکس لیست کلاس ها هم مرتب می شوند. (class0: ایندکس هایی از لیست clips که مقدار Continuous/Sideway Up گرفته اند و class1, class2, class3 به ترتیب برای کلاس های Unknown, Flat, Down می باشند).

```
#-----calculate D_i-----
D_i = []
tmp = []
for i in range(len(class0)): # for class Up
    sum = 0

    xj = clips[class0[i]][0]
    clipp = clips[class0[i]]
    for j in range(len(clipp)):
        sum += math.dist([features.index(clipp[j]) , clipp[j][0]], [features.index(xj), xj[0]])
    tmp.append(1/sum)
if len(tmp) != 0:
    tmp, class0 = zip(*sorted(zip(tmp, class0), reverse=True))
    tmp , class0 = list(tmp), list(class0)
D_i.append(tmp)
```

❖ در مرحله هشتم از بین کلاس ها ، کلاسی را که کمترین مقدار را دارد انتخاب میکنیم و به آن اندازه از همه کلاس ها اولین عناصر را انتخاب می کنیم و به عنوان داده آموزش (train set) در نظر می گیریم. سپس یک کلاس به نام holdClass می سازیم که در آن به ازای کلاس Continuous/Sideway Up مقدار ۱ ، به ازای Continuous/Sideway Down مقدار ۲ ، به ازای Unknown مقدار ۳ و به ازای کلاس Flat مقدار ۴ را قرار می دهیم.

```
holdClass = []
for clas in classes:
    if 'Up' in clas:
        holdClass.append(1.0)
    elif 'Down' in clas:
        holdClass.append(2.0)
    elif 'Unknown' in clas:
        holdClass.append(3.0)
    else:
        holdClass.append(4.0)
```

❖ در مرحله آخر از روش Random Forest پیش بینی می کنیم که کلیپ بعدی متعلق به کدام کلاس می باشد.

```
B = 10
randF = RandomForestRegressor(n_estimators=B)

tree = randF.fit(tmpClip[:550], holdClass2[:550])
pred = tree.predict(tmpClip[550:])
print(metrics.r2_score(holdClass2[550:], pred))
```