**به نام خدا**

در کتابخانه  **scikit-learn**  یا sklearn، می‌توانید از **شاخص جینی (Gini Index)** و **آنتروپی (Entropy)** برای تنظیم معیار تقسیم (splitting criterion) در مدل‌های درخت تصمیم (**Decision Trees**) استفاده کنید. این معیارها تعیین می‌کنند که چگونه ویژگی‌ها برای تقسیم گره‌ها انتخاب شوند.

**تفاوت Gini و Entropy:**

* **Gini Index**:
  + محاسبات سریع‌تر دارد و برای داده‌های بزرگ مناسب‌تر است.
  + مقدار آن بین **۰ تا ۰.۵** تغییر می‌کند (هرچه کمتر، بهتر).
  + فرمول:



که pi*pi*​ نسبت نمونه‌های کلاس i*i* در یک گره است.

* **Entropy**:
  + از نظر تئوری اطلاعات، میزان بی‌نظمی را اندازه‌گیری می‌کند.
  + مقدار آن بین **۰ تا ۱** است (هرچه کمتر، بهتر).
  + فرمول:



### ****نحوه اعمال Gini و Entropy در scikit-learn:****

در sklearn.tree.DecisionTreeClassifier (برای طبقه‌بندی) یا DecisionTreeRegressor (برای رگرسیون)، می‌توانید پارامتر criterion را تنظیم کنید:

#### **۱. استفاده از Gini Index پیش‌فرض در درخت‌های طبقه‌بندی:**

from sklearn.tree import DecisionTreeClassifier

model = DecisionTreeClassifier(criterion='gini') *# معیار: جینی*

model.fit(X\_train, y\_train)

#### **۲. استفاده از Entropy (برای تقسیم بر اساس آنتروپی):**

model = DecisionTreeClassifier(criterion='entropy') *# معیار: آنتروپی*

model.fit(X\_train, y\_train)

#### **۳. نمایش اهمیت ویژگی‌ها بر اساس Gini/Entropy:**

import pandas as pd

*# اهمیت ویژگی‌ها (بر اساس کاهش ناخالصی)*

feature\_importance = pd.DataFrame({

'Feature': X\_train.columns,

'Importance': model.feature\_importances\_

}).sort\_values('Importance', ascending=False)

print(feature\_importance)

| **معیار** | **سرعت** | **کاربرد** | **مقدار** |
| --- | --- | --- | --- |
| **Gini** | سریع‌تر | داده‌های بزرگ | ۰ تا ۰.۵ |
| **Entropy** | کندتر (به دلیل محاسبه لگاریتم) | وقتی تفسیرپذیری مهم است | ۰ تا ۱ |

**نتیجه‌گیری:**

* اگر **سرعت اجرا** مهم است → از criterion='gini' استفاده کنید.
* اگر **دقت و تفسیرپذیری** مدل مهم است → criterion='entropy' را انتخاب کنید.