

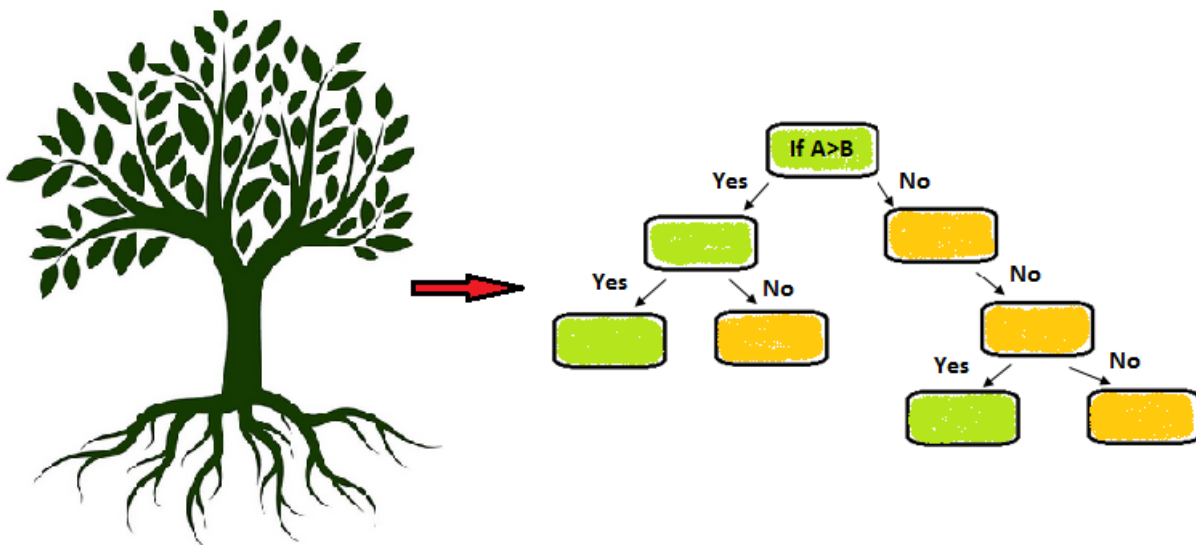
به نام خدا

تمرین درخت تصمیم

درخت تصمیم :

درخت تصمیم یک مدل فلوچارت مانند است که با ترسیم مشاهدات در مورد یک آیتم تا نتیجه‌گیری در مورد متغیر هدف آن به پیش بینی نتایج کمک می‌کند. درخت شامل گره‌هایی است که ویژگی‌های داده‌ها را نشان می‌دهد، شاخه‌هایی که قوانین تصمیم‌گیری را نشان می‌دهند و برگ‌هایی که نشان‌دهنده نتایج یا برچسب‌های کلاس هستند.

به عبارت دیگر، یک الگوریتم درخت تصمیم، یک ساختار درختی از داده‌های آموزشی ایجاد می‌کند که مجموعه‌ای از تصمیمات را نشان می‌دهد که منجر به پیش‌بینی متغیر هدف می‌شود. این الگوریتم با تقسیم بازگشتی داده‌ها به زیرمجموعه‌ها بر اساس ویژگی‌هایی کار می‌کند که بیشترین اطلاعات را تا زمانی که پی بینی یا تصمیم نهایی گرفته شود، فراهم می‌کند.



دیتاست :

این مجموعه داده شامل اطلاعاتی برای تخمین سطوح چاقی در افراد از کشورهای مکزیک، پرو و کلمبیا است که بر اساس عادات تغذیه‌ای و وضعیت فیزیکی آن‌ها تهیه شده است. این داده شامل 17 ویژگی و 2111 رکورد است و هر رکورد با متغیر کلاس NObesity (سطح چاقی) برچسب‌گذاری شده است، که امکان دسته‌بندی داده با استفاده از ارزش‌های وزن کم، وزن نرمال، چاقی سطح I، چاقی سطح II، چاقی نوع I، چاقی نوع II و چاقی نوع III را فراهم می‌کند. 77٪ از داده‌ها به صورت مصنوعی با استفاده از ابزار Weka و فیلتر SMOTE تولید شده‌اند و 23٪ دیگر به صورت مستقیم از کاربران از طریق یک پلتفرم وب جمع‌آوری شده‌اند.

سوال های تمرین :

پیش پردازش:

- (1) ابتدا فایل ObesityDataSet با کتابخانه های پایتون باز کنید. دیتاست را با روش هایی که در درس خوانده اید تمیز کنید و هر مرحله را توضیح دهید.
- (2) تعدادی از ستون ها نیاز به نرمال سازی دارند. آنها را تشخیص دهید و توضیح دهید چه زمانی از این عمل استفاده میکنید.

آموزش مدل:

- (3) با نرخ بیست درصد برای تست و هشتاد درصد برای ترین دیتاست را تقسیم کنید. پس از آن با استفاده از کتابخانه scikit-learn و با استفاده از الگوریتم درخت تصمیم مدل را آموزش دهید.
- امتیازی مثبت : با استفاده از الگوریتم Random Forest مدل را آموزش داده و با نتایج درخت مقایسه کنید.

خروجی:

- (4) نتیجه خروجی را با توجه به معیار های زیر تحلیل کنید.

1. precision

2. recall

3. F1-score

4. Accuracy