

# گزارش کار بخش پیاده‌سازی تمرین سوم - درخت تصمیم

محمد لشکری

۶ دی ۱۴۰۰

متغیرهای مستقل و وابسته به صورت زیر انتخاب شده‌اند:

$X = \{\text{hobby, age, edu\_level, material\_status}\}$

$Y = \{\text{class}\}$

متغیرهای `hobby, edu_level, material_status` به صورت عددی در دیتاست ذخیره شده‌اند. قبل از آموزش نوع آنها به رشته تغییر کرده است.

## ۱ درخت تصمیم

در ابتدا بدون فرایند تنظیم عمق بیشینه<sup>۱</sup> و سپس با تکنیک Cross validation با ۵ فلد و برای مقادیر ۱ تا ۱۰ جهت تنظیم پارامتر عمق بیشینه انجام شد. عمق درخت در حالت اول ۱۰ و در حالت دوم (پس از تنظیم) ۶ به دست آمد. نتایج به صورت زیر قابل مشاهده است:

جدول ۱: دقت طبقه بندی درخت تصمیم

عمق	دقت داده‌های آموزش	دقت داده‌های تست
۱۰	۰/۹۱	۰/۸۶
۶	۰/۸۶	۰/۷۵

در حالت دوم اختلاف دقت آموزش و تست بیشتر است. اما در حالت اول که عمق بیشتری دارد این مقدار کمتر است. پس در حالت دوم underfitting رخ داده است. لازم به ذکر است مدل اول (بدون فرایند تنظیم) تا زمانی که در یک گره، کلاس‌های متفاوت وجود دارد و تعداد نمونه‌ها در هر گره حداقل ۲ است ادامه می‌یابد که با این روش، عمق ۱۰ برای درخت به دست آمد.

<sup>۱</sup>max\_depth hyperparameter tuning