

## پروژه دوم مبانی هوش محاسباتی

### مقدمه

دیتاستی مرتبط با ماشین‌ها در اختیار شما قرار گرفته است؛ دیتاست فوق به دو بخش `train` و `test` تقسیم شده است. در این پروژه از شما می‌خواهیم تا درخت تصمیمی بسازید، که بتواند داده‌های آموزشی را بر اساس مدل ماشین، طبقه‌بندی کند؛ سپس از دیتای `test` برای ارزیابی عملکرد مدل خود استفاده کنید. درخت تصمیم ساخته شده، باید منطبق بر موارد تدریس شده در کلاس باشد.

### ۱ فاز اول

این دیتاست، شامل تعدادی `Missing data` می‌باشد. با توجه به روش‌های مطرح شده در کلاس درس، می‌بایست این مورد را هندل کنید. نتایج حاصل از هر کدام از روش‌ها را بررسی و تحلیل نمائید؛ همچنین نتایج را با هم مقایسه کنید. در این بخش برای `load` کردن و پردازش دیتا، می‌توانید از کتابخانه `Pandas` استفاده نمائید.

### ۲ فاز دوم

همانطور که در درس خوانده اید، از الگوریتم `ID3` برای ساخت درخت تصمیم استفاده می‌شود؛ با استفاده از این الگوریتم درخت تصمیم مناسبی بسازید که با محاسبه `Entropy` و `Information Gain` در هر مرحله `Attribute` مناسب را انتخاب کند. برای این منظور نیاز به پیاده‌سازی کلاس `Node` دارید؛ همچنین توابع `Entropy` و `Information Gain` را نیز خودتان باید پیاده‌سازی کنید. در این فاز موارد زیر را بررسی و گزارش کنید.

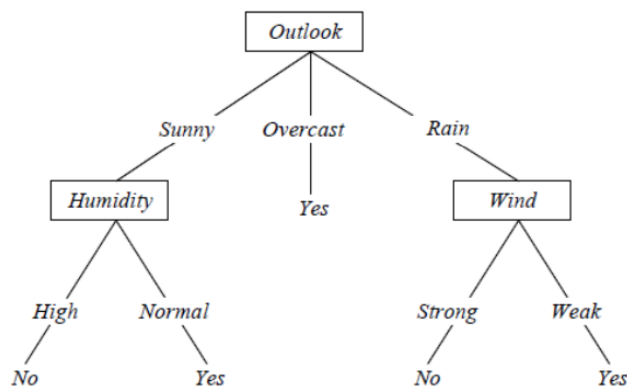
- نحوه تشکیل درخت و انتخاب فیچرها را به طور کامل بررسی کنید.
- راه حل خود را برای فیچرهای پیوسته توضیح دهید (دقت داشته باشید که باید مقدار بهینه `Threshold` را برای گسسته‌سازی این نوع فیچر ها استفاده نمائید).
- بررسی کنید که چه زمانی `Entropy` و `Information Gain` به مقدار حداکثر خود می‌رسند.

### ۳ فاز سوم

یکی از مشکلات درخت تصمیم، `Overfit` شدن آن است؛ بررسی کنید که این مشکل چه زمانی به وجود می‌آید و برای حل این مشکل حداقل دو راه حل را بررسی نمائید (نیازی به پیاده‌سازی نیست).

### ۴ فاز چهارم

در نهایت درخت ساخته شده را `visualize` کنید (مشابه شکل ۱). در هر `Node` درخت، نام `Attribute` ها، `Entropy` و `Information Gain` مربوطه را ذکر کنید. برای این منظور می‌توانید از کتابخانه‌هایی مانند `Plotly`، `Graphviz` و یا ابزارهای مشابه استفاده نمائید.



شکل ۱: درخت تصمیم

## توضیحات تکمیلی

- انجام پروژه می‌تواند در قالب گروه‌های دو نفره و یا به صورت انفرادی صورت گیرد.
- علاوه بر سورس کد پروژه، فایل مستندات نیز باید آپلود شود.
- در فایل مستندات پروژه نام هر دو عضو گروه را ذکر کنید و آپلود فایل‌ها همین که توسط یکی از اعضای گروه انجام شود کافی است.
- هر گونه شباهت نامتعارف بین کد شما و کد سایر گروه‌ها و یا کدهای موجود بر روی اینترنت تقلب محسوب می‌شود و نمره‌ای برای این پروژه دریافت نخواهید کرد.
- در صورت نوشتن داکيومنت تمیز (برای مثال با  $\text{LaTeX}$ ) نمره اضافه برای شما در نظر گرفته خواهد شد.
- استفاده از هرگونه روش خلاقانه نمره اضافی خواهد داشت.
- استفاده از کتابخانه‌ها و فریم‌ورک‌های آماده به جز مواردی که در صورت پروژه از شما خواسته شده تا پیاده‌سازی کنید، بلامانع است.
- فایل شامل سورس کد پروژه و مستندات را در قالب فایل zip و با نام شماره دانشجویی خود ذخیره و ارسال نمایید.
- در صورت داشتن هرگونه سوال می‌توانید با [kourosh\\_hsz](#) و یا [mhmdrzrs](#) در ارتباط باشید و یا در گروه درسی مطرح نمایید.

با احترام - تیم حل تمرین