

تمرین دوم

۱) مطالعه دو مقاله علمی-پژوهشی به عنوان بخشی از این تمرین برای شما در نظر گرفته شده است که منبع مطالعاتی خوبی برای یادآوری مباحث تدریس شده و مشاهده کاربرد عملی آن در تحقیقات کاربردی نیز می‌باشد. یک مقاله با عنوان "افزایش نرخ نفوذ به شبکه‌های کامپیوتری با استفاده از درخت تصمیم" و دیگری با عنوان "استخراج دانش از داده‌های بیماران دیابتی با استفاده از درخت تصمیم C5.0" در نظر گرفته شده است. این مقالات توسط محققان ایرانی در حوزه داده کاوی به صورت خیلی روان به زبان فارسی به نگارش درآمده است. پس از مطالعه‌ی مقالات موردنظر، خلاصه‌ی مناسب و مفیدی از برداشت خود از هر کدام از این مقالات را در قالب یک پاراگراف (شامل حداقل ۲۵۰ کلمه و حداکثر ۵۰۰ کلمه) ذکر کنید.

۲) با استفاده از مطالعه منابع درسی به سوالات زیر پاسخ دهید.

۱-۲) دسته‌بندی داده‌ای و خوشه‌بندی داده‌ای را با یکدیگر مقایسه کنید.

۲-۲) نام چند روش معروف برای دسته‌بندی داده‌ای را نام ببرید.

۳-۲) نقطه قوت و ضعف روش دسته‌بندی مبتنی بر قانون چیست؟

۴-۲) برای ساخت درخت تصمیم به صورت حریصانه برای اولویت‌دهی جهت انتخاب نودها چه معیارهایی وجود دارد؟

جدول ۱

	X ₁	X ₂	X ₃	C
A	3	1	10	Y
B	2	1	20	N
C	3	1	30	Y
D	3	0	10	Y
E	3	1	20	Y
F	2	1	30	N
G	3	0	30	Y
H	2	0	10	N

۳) با توجه مجموعه دادگان به شرح جدول ۱ (ستون‌های X₁, X₂, X₃ ویژگی و ستون C برچسب تصمیم)، مطلوبست

۱-۳) اگر از درخت تصمیم برای دسته‌بندی استفاده شود، براساس معیار بهره اطلاعاتی (information gain)

کدامیک از ویژگی‌های X₁, X₂, X₃ در ریشه درخت قرار می‌گیرد؟

۲-۳) اگر با استفاده از یک الگوریتم دسته‌بندی دلخواه، نمونه‌ها به صورت $Y = \{A, B, C, D, E\}$ و $N = \{F, G, H\}$

شناسایی شوند، معیارهای ارزیابی Precision, Recall, F-measure, Accuracy را محاسبه کنید.

به نکات زیر قبل از پاسخ دهی و آماده‌سازی پاسخ‌های خود دقت کنید.

- برای پیدا کردن پاسخ‌های خود، مطالعه و استفاده از هرگونه منابع مطالعاتی در صورتی که مطابق با سرفصل درس باشد، مجاز است.
- هدف از این تمرین، مطالعه و یادگیری مباحث درسی در طول ترم است، تقلب یا کپی علاوه بر آنکه فرصت یادگیری را از شما سلب می‌کند، ممکن است، باعث کسر نمره تمرین شود.
- آماده‌سازی و ارسال تمرینات به صورت انفرادی است.
- پاسخ‌های خود را به صورت فیزیکی (دست‌نویس بر روی برگه کاغذی) به طور خوانا و منظم آماده کنید.
- آماده‌سازی تمرین به صورت تایپی قابل قبول نیست.
- از آماده‌سازی پاسخ‌های خود به صورت تک‌رو پرهیز کنید و از هر دو روی برگه استفاده کنید.
- در کلیه برگه‌ها، نام، نام خانوادگی، شماره دانشجویی و شماره صفحه را درج نمایید.

- کلیه مراحل محاسباتی خود را ذکر کنید. (ذکر صرفاً پاسخ نهایی قابل قبول نیست)
- مهلت تحویل تمرین حداکثر تا روز امتحان میانترم است. (غیر قابل تمدید).
- طرح سوال در امتحان میانترم از مقالات محتمل است.
- به ازای هر لحظه تاخیر پس از مهلت مقرر تا ۲۴ ساعت اول از مهلت مقرر ۳۰٪ از نمره، از ۲۴ ساعت تا ۴۸ ساعت تاخیر از مهلت مقرر ۶۰٪ از نمره و پس از ۷۲ ساعت از مهلت مقرر ۱۰۰٪ از نمره تمرین به عنوان جریمه بی‌نظمی کاسته می‌شود.

بازبینی ۱۲ آبان ۱۳۹۶