بەنام پروردگار

مبانی داده‌کاوی

نیم‌سال اول ۱۳۹۸-۱۳۹۷

مدرس: سید مهدی وحیدی پور

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 داده‌کاوی به صورت خیلی روان به زبان فارسی به نگارش درآمده است. پس از مطالعه‌ی مقالات موردنظر، به سوالات زیر در مورد هر دو مقاله جواب دهید:

الف. مساله مقاله چیست؟

ب. از چه روشی استفاده کرده است؟

ج. از چه داده هایی استفاده کرده است؟ ویژگیهای آن چیست؟

د. روش پیشنهادی خود را با چه روشهای دیگری مقایسه کرده است؟ معیار مقایسه چیست؟

ه. برتری روشهای ارایه شده در مقاله نسبت به سایر روشهای مشابه چیست؟

۲) با توجه مجموعه دادگان به شرح جدول ۱ (ستون‌های X_1, X_2, X_3 ویژگی و ستون C برچسب تصمیم)، مطلوبست

۲-۱) اگر از درخت تصمیم برای دسته‌بندی استفاده شود، براساس معیار بهره اطلاعاتی (information gain) کدامیک از ویژگی‌های X_1, X_2, X_3 در ریشه درخت قرار می‌گیرد؟

۲-۲) 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.

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

زمان تعریف: ۱۰ آبان ۹۷، مدت زمان تحویل: ۵ آذر ۹۷

نحوه پاسخگویی: دست نویس و تحویل در کلاس

بازبینی ۹ آبان ۱۳۹۷