

## کوییز درس داده کاوی، ۶ خرداد ۱۳۹۹

لطفاً، دانشجویان با شماره دانشجویی فرد سوال ۱ و دانشجویان با شماره دانشجویی زوج سوال ۲ را حل کرده و پاسخهای خود را به شماره 09369084958 واتس آپ نمایند. نامگذاری فایل‌های ارسالی باید به صورت `lastName-ID.ext` باشد، به طوریکه `lastName` به نام خانوادگی دانشجو، `ID` به شماره دانشجویی و `ext` به پسوند فایل ارسالی اشاره می‌نماید. مثال: `kazemi-9731063.jpg`

\* در صورت بروز هرگونه مشکل فنی، دانشجویان پاسخهای خود را به آدرس `nazerfard@yahoo.com` ایمیل نمایند.

\*\* رعایت نکردن موارد فوق باعث کسر نمره خواهد شد.

\*\*\* در صورت کشف هرگونه تقلب، تمامی دانشجویانی که در این امر مشارکت داشته اند، در درس مردود خواهند شد.

**سوال (۱)** یک نمایشگاه ماشین، به منظور تعیین اینکه آیا یک ماشین مناسب استفاده یک خانواده است یا خیر، ویژگیهای زیر را در نظر گرفته است:

Features/attributes	Possible values
Price	Economy, Not-economy
Engine Speed	Low, Medium, High
SUV	Yes, No
Family car	Yes, No

فرض کنید مجموعه داده‌های زیر قابل استناد به عنوان داده‌های آموزشی هستند:

Price	Engine Speed	SUV	Family Car
Economy	High	Yes	No
Economy	Low	No	No
Economy	Medium	Yes	Yes
Not-economy	Medium	No	Yes
Not-economy	High	Yes	No
Not-economy	Medium	Yes	Yes

در صورتیکه تصمیم گرفته باشید از درخت تصمیم به منظور پیشنهاد یک ماشین بعنوان ماشین خانوادگی استفاده کنید، به سوالات طرح شده پاسخ دهید:

الف. (۱ نمره) آنترپی (entropy) مجموعه داده‌های آموزش را محاسبه کنید. محاسبات خود را نشان دهید.

ب. (۳ نمره) درخت تصمیم نهایی یاد گرفته شده توسط الگوریتم ID3 را نمایش دهید. دلیل انتخاب الگوریتم را کاملاً توجیه نمایید.

ج. (۱ نمره) درخت تصمیمی که با استفاده از الگوریتم C4.5 حاصل میشود را نمایش دهید. محاسبات خود را نشان دهید.

**سوال ۲)** تصور کنید که شما مسوول تایید یا رد تقاضای اعطای وام در یک بانک هستید. هیات مدیره بانک تصمیم گرفته به منظور تایید یا رد تقاضای مشتریان ویژگیهای زیر را در نظر بگیرد:

Features/attributes	Possible values
Income	Low, Medium, High
Education	HS, BS, MS, PhD
Debt	Low, Medium, High
Decision	Approve, Reject

در نظر بگیرید که داده‌های زیر از درخواستهای قبل موجود و مورد تایید هیات مدیره میباشد:

Income	Education	Debt	Decision
Low	HS	High	Approve
Low	BS	High	Approve
Low	MS	High	Approve
Medium	MS	Medium	Reject
High	MS	Medium	Reject
High	PhD	Low	Reject

در صورتیکه تصمیم گرفته باشید از درخت تصمیم به منظور تصمیم گیری در خصوص تایید یا رد تقاضای اعطای وام استفاده کنید، به سوالات طرح شده زیر پاسخ دهید:

الف. (۱ نمره) آنترپی (entropy) مجموعه داده‌های آموزش را محاسبه کنید. محاسبات خود را نشان دهید.

ب. (۳ نمره) درخت تصمیم نهایی یاد گرفته شده توسط الگوریتم ID3 را نمایش دهید. دلیل انتخاب الگوریتم را کاملاً توجیه نمایید.

ج. (۱ نمره) درخت تصمیمی که با استفاده از الگوریتم C4.5 حاصل میشود را نمایش دهید. محاسبات خود را نشان دهید.