

هدف از این پروژه آشنایی عملی با برخی از امکانات ابزارهای داده کاوی مانند Rapid miner است. خروجی این پروژه یک گزارش کامل است. در ابتدای این گزارش باید مجموعه داده مربوط به خودتان را توضیح بدهید. برای یافتن اسم داده مربوط به خودتان، از جدول زیر استفاده کنید. داده‌ها را از آدرس <https://archive.ics.uci.edu/ml/datasets.html> دانلود کنید.

نام داده	ویژگی شماره دانشجویی	نام داده	ویژگی شماره دانشجویی
Iris	عدد سمت راست ۰، ۱، ۴	Glass Identification	عدد سمت راست ۷، ۹
Badges	عدد سمت راست ۳، ۲	Ionosphere	عدد سمت راست ۵، ۶، ۸

در ادامه گزارش باید سه سوال زیر را پاسخ دهید (جداول خواسته شده را پر کنید!). دقت کنید که تنها پر کردن جداول نمره‌ای نخواهد داشت. برای گرفتن نمره کامل باید برای هر سوال، گزارش انجام کار را تهیه کنید: روند انجام تمرین در ابزار، تصاویر انجام مراحل مختلف، نحوه محاسبه مقادیر، تنظیمات و پارامترهای هر روش، هر فرضی که دانشجو برای انجام الگوریتم داشته است، هر توضیحی که برای فهم روش نیاز است، ...

متن ساده و روان گزارش نمره بالاتری خواهد داشت. دقت کنید که تمام تصاویر و جداول گزارش ارسالی باید شماره یا برجسب (Caption) داشته باشند؛ مانند جدول ۱ یا شکل ۳. از فونت‌های استاندارد استفاده کنید. برای آشنایی با نحوه صحیح نگارش می‌توانید به فایل "ارائه چهارم: ارائه نوشتاری" در درس شیوه پژوهش و ارائه اینجانب مراجعه کنید^۱.

(۱) تاثیر Cross validation 10-folds بر کارایی Decision tree: بر اساس مجموعه داده مربوط به خود، جدول زیر را کامل کنید.

		TP Rate	FP Rate	Precision	Recall	F-measure
1	Decision tree with use training set					
2	Decision tree with Cross-validation					

(۲) بررسی روشهای Ensemble

		TP Rate	FP Rate	Precision	Recall	F-measure
1	Ada boost					
2	Random Forest					

(۳) داده‌های خود را بدون برجسب کلاس در نظر بگیرید و آنها را خوشه‌بندی کنید. حال بر اساس برجسب موجود کلاس درصد نمونه‌های اشتباه را به ازای مقدار مختلف تعداد خوشه ۲، ۳، ۴ و ۵ در جدول زیر گزارش کنید.

Method \ K	2	3	4	5
K-Means				
K-Medoids				

- مهلت تحویل تمرین حداکثر تا ساعت ۲۴:۰۰ مورخه ۲۰ بهمن ۱۳۹۶ می‌باشد (غیر قابل تمدید). 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.

^۱ <https://faculty.kashanu.ac.ir/vahidipour/fa/page/14142/> شیوه‌ارایه و پژوهش