

تاریخ تحویل ۳۰ دی ۱۳۹۷

در این پروژه قصد داریم از مجموعه داده‌ی Mushroom در آدرس <https://archive.ics.uci.edu/ml/datasets/Mushroom> استفاده کنیم.

- می‌توانید از ابزارهای داده کاوی نظیر Rapidminer و Weka استفاده کنید.
- **توجه: حل پروژه با پایتون نمره اضافه دارد.**
- گزارش کاملی از آنچه انجام می‌دهید با توضیحات کافی در قالب pdf به آدرس db.laboratory2014@yahoo.com ارسال کنید.

الف) توضیح کاملی در مورد دیتاست، ویژگی‌های آن، ویژگی کلاس این دیتاست و ... ارائه دهید.

توجه: در سوال ب تا د جهت تقسیم دیتاست به دو دسته‌ی آزمایش و آموزش از روش k-fold cross validation استفاده کنید. K را می‌توانید عددی بین ۵ تا ۱۰ انتخاب کنید.

ب) درخت‌های تصمیم ID3 و CART را برای این دیتاست ایجاد کنید. قبل از اعمال الگوریتم‌ها در صورت نیاز دیتاست را پیش-پردازش کنید.

ج) قوانین را از درخت تصمیم ID3 یا CART استخراج کنید.

د) کدام یک از درخت‌ها دسته‌بندی بهتری روی این دیتاست انجام می‌دهد؟ (برای مقایسه مدل‌ها از سه معیار Precision, recall و F-measure استفاده کنید)

ه) از روش KNN جهت کلاس‌بندی داده‌ها استفاده کنید. در این سوال جهت تقسیم دیتاست به دو دسته‌ی آزمایش و آموزش به روش hold out عمل کنید. دقت و صحت مدل را با سه معیار Precision, recall و F-measure مقایسه کنید.

موفق باشید

میترا عیسایی