

دوره یادگیری ماشین نظارت شده  
تمرین دوم: k نزدیکترین همسایه، درخت تصمیم و معیارهای ارزیابی

✚ شرح تمرین

- یک مجموعه داده به نام *BOM.csv* مربوط به داده های آب و هواشناسی در پوشه دیتاست موجود است.
- الف) مجموعه داده را خوانده و پیش پردازش های لازم را اعمال نموده و آن را به دو مجموعه داده ای آزمون و آموزش به نسبت ۲۰:۸۰ تقسیم نمایید.
- ب) یک مدل *K-NN* برای پیش بینی بارش فردا (*Rain Tomorrow*) آموزش داده و بهترین *K* را بر اساس معیار *F1* بیابید.
- ج) برای پیش بینی بارش فردا به صورت *real time* در کوتاه ترین زمان ممکن کدام یک از این دو الگوریتم، درخت تصمیم و *KNN* مناسب تر است؟ چرا؟