

شرح تمرين

- هدف از دومین تمرين اين دوره، گرد هم آوردن تمام مهارت های لازم برای انجام سه مرحله اول یک پروژه یادگیری ماشین (Data ingestion, Data Validation, Data Transformation) است.
- مجموعه داده در نظر گفته شده برای اين تمرين را می توانيد از آدرس زير دانلود نمایيد. از اين مجموعه داده برای آموزش مدل پيش بينی نوع پوشش جنگل بر اساس متغیر های ارائه شده استفاده کنيد.

<https://archive.ics.uci.edu/dataset/31/covertype>

- ابتدا تعداد ویژگی های مجموعه داده را با حداقل سه روش متفاوت کاهش دهيد (Feature Selection). اين کار به کاهش پیچیدگی مدل شما و صرفه جویی در منابع در حین آموزش کمک می کند. اين نکته را در نظر بگيريد که کاهش ویژگی ها نباید بيشتر از ۵ درصد از دقت مدل را کاهش دهد.
- سپس با آماده شدن زير مجموعه انتخاب شده از ویژگي ها، پايپ لайн ديتا که شامل validating, ingesting, transform مجموعه داده می باشد را طراحی کنيد. انجام مراحل زير در طراحی پايپ لайн ضروري می باشد.

ExampleGen	<input type="radio"/>
StatisticsGen	<input type="radio"/>
SchemaGen	<input type="radio"/>
ExampleValidator	<input type="radio"/>
Preprocessing function	<input type="radio"/>
Transform	<input type="radio"/>