



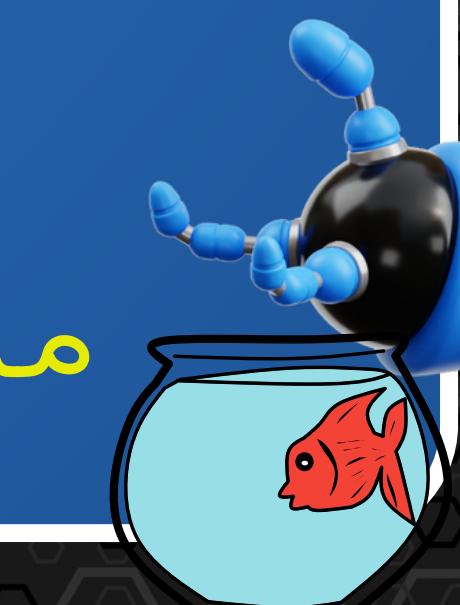
تمرين 6:
يک نوروزي

مهلت:

1403/01/15

نمره کل تمرین: 1200
استاد: دکتر مسعود کاویانی

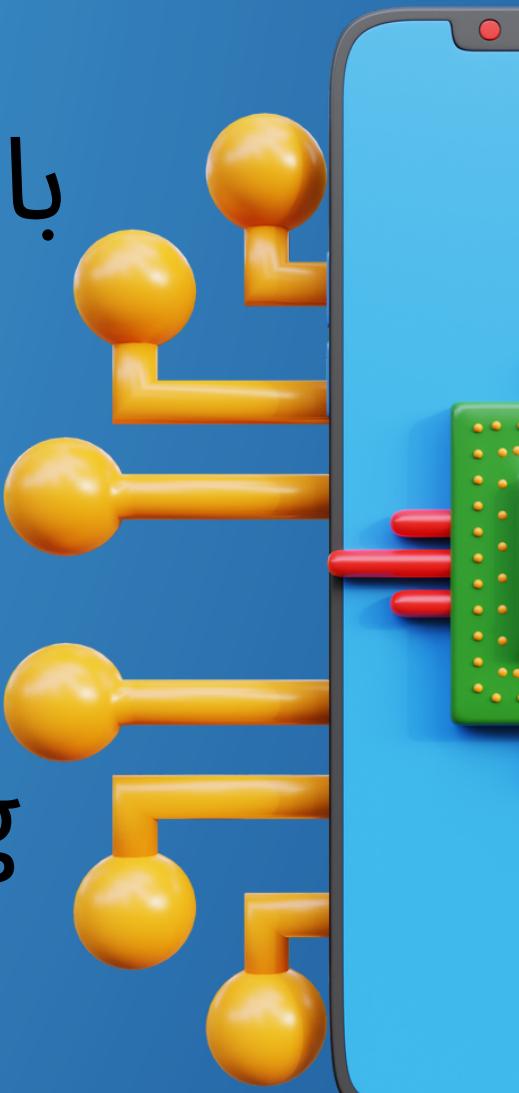
مدرسء متخصص دانشمند داده



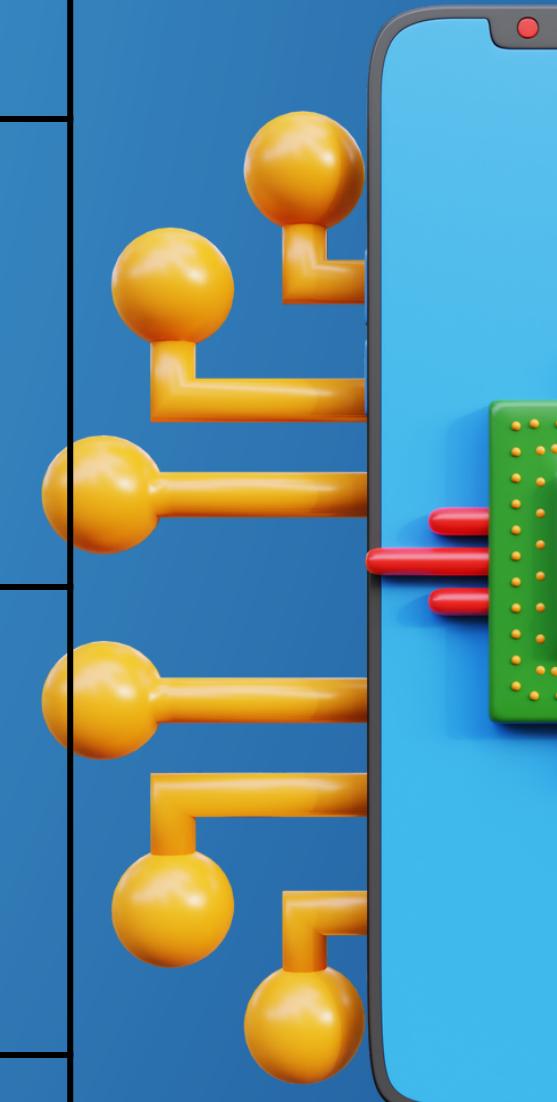
مجموعه داده‌ی ضمیمه (bank.csv) را دانلود کنید. این مجموعه داده در مورد یک کسب و کار است. هر سطر یک مشتری را نشان می‌دهد و هر ستون یک ویژگی از آن مشتریست. ستون لا هم تارگت است.

با استفاده از اطلاعات زیر (که هر ستون را توضیح داده است) ابتدا بر روی داده‌ها به صورت کامل EDA انجام دهید. سعی کنید با استفاده از نمودارهای مختلف، داده‌ها را از زوایایی متفاوت بررسی کنید.

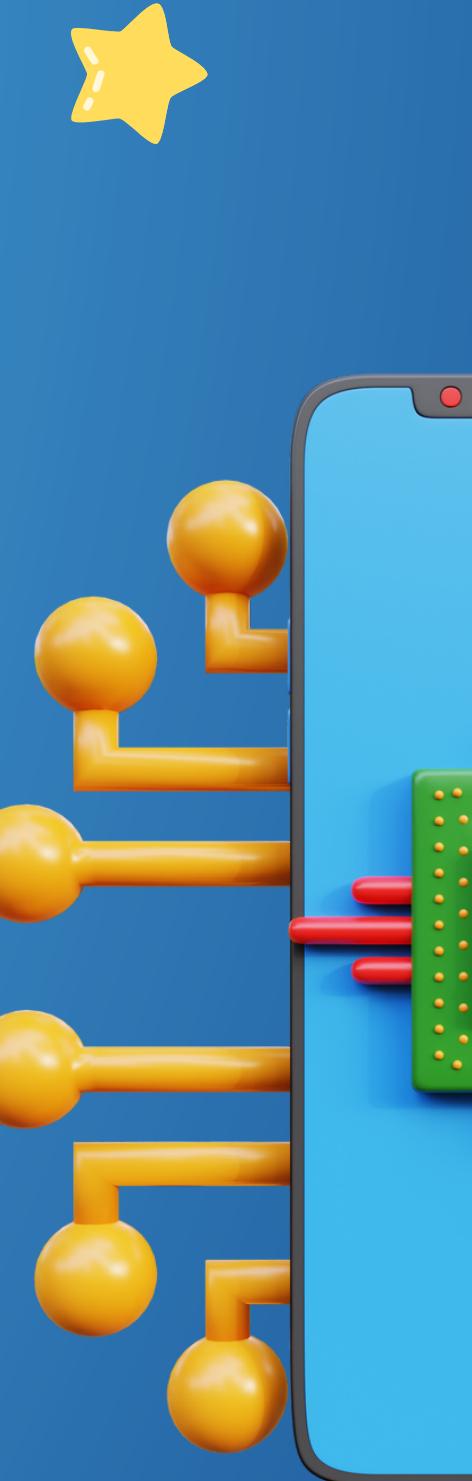
سپس با استفاده از روش‌های پیش‌پردازش داده، عملیات preprocessing را بر روی این داده‌ها انجام دهید. سپس داده‌ها را به دو قسمت train/test تقسیم کنید. بعد از آن با استفاده از حداقل سه الگوریتم یادگیری ماشین (مثل SVM، KNN و...) یادگیری را بر روی مجموعه‌ی آموزشی انجام داده و سپس بر روی مجموعه‌ی آزمایشی آن را تست کنید.



Variable Name	Type	Description
age	Integer	
job	Categorical	<p>type of job (categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')</p>
marital	Categorical	<p>marital status (categorical: 'divorced', 'married', 'single', 'unknown'; note: 'divorced' means divorced or widowed)</p>
education	Categorical	<p>(categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')</p>

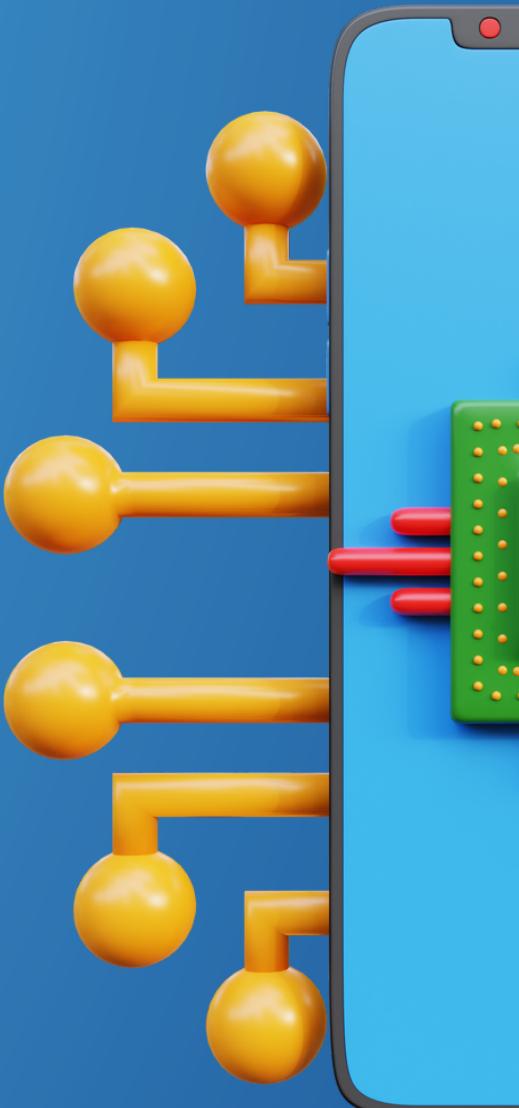


default	Binary	has credit in default?
balance	Integer	average yearly balance
housing	Binary	has housing loan?
loan	Binary	has personal loan?
contact	Categorical	contact communication type (categorical: 'cellular','telephone')





day_of_week	Date	last contact day of the week
month	Date	last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
duration	Integer	last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.
campaign	Integer	number of contacts performed during this campaign and for this client (numeric, includes last contact)





pdays	Integer	number of days that passed by after the client was last contacted from a previous campaign (numeric; -1 means client was not previously contacted)
previous	Integer	number of contacts performed before this campaign and for this client
poutcome	Categorical	outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')
y	Binary	has the client subscribed a term deposit?

