

Step1:

1 داده شما برچسب دار است یا خیر؟ (unsupervised / supervised)

تارگت ما در این مسئله قیمت هر خانه هست. با توجه به اینکه در داده ورودی این ستون برای هر رکورد وجود دارد، پس داده برچسب دار محسوب میشود و میتوان از الگوریتم های supervised برای آن استفاده کرد.

2 این سوال جز کدام یک از وظایف عمده (task) داده کاوی است؟

Prediction یا پیش بینی. که برای پیش بینی مقادیر ناشناخته یا آینده مجموعه داده های مورد علاقه دیگر مفید است. با توجه به اینکه target مقدار عددی و پیوسته هست، پس رگرسیون محسوب میشود.

3 آیا سیستم ما Batch است یا online ؟

A model for predicting housing prices in different regions of California can be considered either a **batch** data mining system or an **online** data mining system, depending on how the data is used in the model.

- In **batch data mining**, data is collected over a period of time and then the model is trained using this collected data. This method is suitable when the data changes infrequently or computational resources are limited.
- In **online data mining**, the model updates itself over time as new data comes in. This method is suitable when there is a need to respond to changes in the data in real time.

So, if my model is trained using data collected over a specific period of time and then used for prediction, it operates in a **batch** mode. However, if your model updates itself continuously as new data comes in, it operates in an **online** mode.

Since I have a training data set and I do not intend to use other data in the future, I present a batch model.

Step2:

<https://colab.research.google.com/drive/1Qdwhuzm1WwE8MLgmjEqcAloA7aEDlvd8?usp=sharing>