

## پروژه دوم و سوم : تصمیم گیری – درس یادگیری ماشین

هدف : طراحی الگوریتم های تصمیم گیری مبتنی بر یکی از " مجموعه داده ها" زیر

پروژه اختیاری: الگوریتم خطی مبتنی بر رگرسیون لجستیک

پروژه دوم: الگوریتم SVM خطی

پروژه سوم: الگوریتم kNN

گام های مورد نیاز:

۱. Import نمودن کتابخانه های مورد نیاز
  ۲. بارگذاری و خواندن دیتاست با استفاده از کتابخانه
  ۳. اکتشاف دیتاست و بررسی داده به صورت کامل
  ۴. شناسایی داده های گم شده و Missing Value
  ۵. ترسیم نمودارهای مختلف جهت درک بهتر از داده ها
  ۶. آموزش مدل با استفاده از رگرسیون خطی چند متغیره
  ۷. معیارهای ارزیابی:
- معیارهای ارزیابی مشخصی تعیین کنید:
۱. MSE (Mean Squared Error)
  ۲. RMSE (Root Mean Squared Error)
  ۳.  $R^2$  Score
  ۴. MAE (Mean Absolute Error)

گزارش ( فشرده شده) : شامل موارد زیر

- فایل های دیتا ست
  - متن گزارش شامل توضیحات مربوط به:
۱. "مجموعه دیتا"
  ۲. ویژگی های مدل

- ۳. متن برنامه
- ۴. نتایج
- ۵. بررسی و تجزیه و تحلیل نتایج
- فایل اجرایی برنامه

## لینک های دیتاست های درس یادگیری ماشین بخش تصمیم گیری

1. <https://www.kaggle.com/datasets/fedesoriano/stellar-classification-dataset-sdss17>

مجموعه داده های دسته بندی ستاره ای

---

2. <https://www.kaggle.com/datasets/datascientistanna/customers-dataset>

خرید اطلاعات مشتری

---

3. <https://www.kaggle.com/datasets/rkiattisak/student-performance-in-mathematics>

پیش بینی عملکرد دانش آموزان

این مجموعه داده حاوی اطلاعاتی در مورد عملکرد دانش آموزان در ریاضیات است

---

4. <https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification>

طبقه بندی قیمت موبایل

---

5. <https://www.kaggle.com/datasets/anandaram/apartment-cost-in-new-york-city>

هزینه مسکن در نیویورک

---

6. <https://www.kaggle.com/datasets/mirzahasnine/loan-data-set>

مجموعه داده های وام

مجموعه داده های تایید وام

---

7. <https://www.kaggle.com/datasets/arjunbasandrai/25-indian-bird-species-with-226k-images>

52 گونه پرنده هندی با ۵۵۲۲ هزار تصویر

مجموعه داده ۵۲ گونه پرنده برای طبقه بندی تصویر

---

8. <https://www.kaggle.com/datasets/fatihb/citibike-sampled-data-2013-2017>

داده های نمونه (5102-5102 Citibike)

---

9. <https://www.kaggle.com/datasets/gpiosenka/cats-in-the-wild-image-classification>

گره بزرگ طبیعت - طبقه بندی تصویر

01 کلاس ۵۲۲۲ قطار، ۲۱ تست، ۲۱ فایل اعتبار سنجی ۵۵۲ X 2 X 552 فرمت jpg

10. <https://www.kaggle.com/datasets/subhajournal/android-malware-detection>

تشخیص بدافزار اندروید

تشخیص بدافزار اندروید با استفاده از یادگیری ماشینی

---

11. <https://www.kaggle.com/datasets/nikhil25803/github-dataset>

---

12. <https://www.kaggle.com/datasets/ulrikthgepedersen/mushroom-attributes>

ویژگی های قارچ

آیا می توانید پیش بینی کنید که قارچ سمی است یا خوراکی؟

---

13. <https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>

مجموعه داده های تجزیه و تحلیل و پیش بینی حمله قلبی

مجموعه داده ای برای طبقه بندی حمله قلبی

---

14. [https://scikit-learn.org/stable/datasets/toy\\_dataset.html](https://scikit-learn.org/stable/datasets/toy_dataset.html)

14-1. `load_iris(*[, return_X_y, as_frame])`

Load and return the iris dataset (classification).

14-2. `load_digits(*[, n_class, return_X_y, as_frame])`

Load and return the digits dataset (classification).

14-3. `load_wine(*[, return_X_y, as_frame])`

Load and return the wine dataset (classification).

14-4. `load_breast_cancer(*[, return_X_y, as_frame])`

Load and return the breast cancer wisconsin dataset (classification)

---