



Home Work

Data science

HW_L04_01_ML

1. title فایل تمرین به صورت (نام تمرین+نام و نام خانوادگی) به انگلیسی باشد.
2. با استفاده از دیتابیس loan موارد خواسته شده را پیاده سازی نمایید.
3. کدهای خود را در یک فایل ipynb، به همراه نتایج و نتیجه گیری ها و توضیحات کامل (mark down) در پنل خود بارگذاری کنید.
4. در صورتی که سوال و یا ابهامی دارید در گروه چت تلگرامی بپرسید.

1. پیش بینی واجد شرایط وام بودن

هدف : خودکار کردن فرآیند بررسی واجد شرایط وام بودن افراد بر اساس اطلاعات صاحب حساب. در این تمرین شما ابتدا داده ها را پیش پردازش می کنید، سپس مدل های مختلف را آموزش می دهید و در نهایت عملکرد آنها را ارزیابی می نمایید.

بخش ۱: پیش پردازش داده ها

(a) **مدیریت مقادیر ناموجود:** برخی ستون های دادگان این تمرین مقادیر نا موجود و گم شده دارند. در این بخش نیاز است بار روش هایی که در این فصل یاد گرفته اید مشکل داده های از دست رفته را برطرف کنید.

(b) **کد گذاری متغیرهای غیر عددی:** مقادیر برخی ستون‌های این دادگان غیر عددی هستند. با توجه به اینکه مدل‌های یادگیری ماشین امکان پردازش مقادیر غیر عددی را ندارند نیاز است در این بخش این مشکل را نیز با استفاده از دانش خود و یا جستجو در اینترنت برطرف کنید.

(c) **تقسیم داده‌های آموزش و تست :** همانطور که می‌دانید ارزیابی مدل‌های یادگیری ماشین نمی‌تواند بر روی داده‌هایی که برای آموزش مدل استفاده شده‌اند انجام گیرد. در این بخش نیاز است داده‌ها را به دو بخش داده‌های آموزش و داده‌های ارزیابی تقسیم کنید.

(d) **نرمال سازی داده‌های عددی :** نقش نرمال سازی داده‌های عددی در یادگیری مدل را توضیح دهید و مدل‌های بخش قبل را یک بار بدون نرمال سازی داده‌ها و یک بار با نرمال سازی داده‌ها آموزش دهید و نتیجه را گزارش نمایید.

بخش ۲: آموزش مدل

(a) مدل‌های یادگیری ماشین زیر را بر روی داده‌های آماده شده از بخش قبل آموزش داده و با تنظیم هایپرپارامترها تلاش کنید بهترین عملکرد هر مدل را بیابید و عملکرد هر یک را بر اساس معیارهایی که تا به اینجا یاد گرفته اید بر روی داده‌های ارزیابی که در بخش قبل جدا نمودید گزارش کنید.

- Logistic regression
- KNN
- ANN

بخش ۳: مقایسه مدلها

(a) حال عملکرد مدل‌های بالا را بر روی دادگان test که در کنار دادگان loan قرار گرفته است امتحان کنید و بحث کنید که کدام مدل در نهایت عملکرد بهتری داشته و دلایل احتمالی برای عملکرد مشاهده شده را ارائه دهید.

بخش ۴: کاوش و بهبود

(a) برای کسب نمره اضافه سعی کنید یک یا دو روش برای بهبود نتایج پیشنهاد و سپس آن‌ها را پیاده سازی کنید.

* مجموعه داده‌ای که در اختیار شما قرار گرفته است، شامل ویژگی‌های زیر است:

جنسیت:	مرد/زن
وضعیت تاهل:	بله/خیر
تحت تکفل:	تعداد افراد تحت تکفل (۰ یا ۱ یا ۲ یا ۳+)
تحصیلات:	فارغ‌التحصیل/غیر فارغ‌التحصیل
خویش فرما:	بله/خیر
درآمد متقاضی:	مقدار عددی درآمد
درآمد همسر متقاضی:	مقدار عددی درآمد
مبلغ وام:	مبلغ وام درخواستی
مدت بازپرداخت وام:	دوره بازپرداخت وام به روز
سابقه اعتباری:	سابقه بازپرداخت موفق وام در گذشته (1: بله، 0: خیر)
منطقه ملک درخواست کننده:	شهری/نیمه‌شهری/روستایی
نتیجه درخواست وام:	وام تایید شده یا نه (Y/N)