



Home Work

Data Science

DataScience_Project

1. با استفاده از دیتاست loan موارد خواسته شده را پیاده سازی نمایید. کدهای خود را در یک فایل ipynb. به همراه نتایج، نتیجه گیری ها و توضیحات کامل (markdown) در یک فایل zip با عنوان شماره تمرین و نام خودتان "YourName_Project.zip" قرار داده و تا تاریخ ۴ خرداد ماه در پنل خود در سایت بارگذاری کنید (عدم ثبت نام شما در نام فایل می تواند باعث اشتباه در ثبت نمره و فیدبک شود).

2. در صورتی که سوال و یا ابهامی دارید در گروه چت تلگرامی بپرسید.

۱. پیش بینی واجدین شرایط وام

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

بخش ۱: پیش پردازش داده ها

(a) مدیریت مقادیر ناموجود: برخی ستون های دادگان این پروژه مقادیر ناموجود و گم شده دارند. در این بخش نیاز است با روش هایی که در این دوره 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 داده‌ها جهت آموزش و ارزیابی:** 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 سه بخش داده‌های آموزش و ارزیابی و تست (۶۵-۲۰-۱۵) تقسیم کنید.

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

بخش ۲: آموزش مدل

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 در این دوره یاد گرفته‌اید بر روی داده‌های ارزیابی که در بخش قبل جدا نمودید گزارش کنید.

- **Logistic regression**

- **KNN**

- **ANN**

بخش ۳: مقایسه مدل‌ها

حال عملکرد مدل‌های بخش قبل را بر روی دادگان تست که در بخش قبل تفکیک نمودید 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

برای کسب نمره اضافه سعی کنید بر اساس آنچه آموخته‌اید یا جست‌وجو در وب 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+)
تحصیالت:	فارغ التحصیل/غیر فارغ التحصیل
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: (۱ بله، ۰ خیر)
منطقه ملک درخواست کننده:	شهری/نیمه شهری/روستایی
نتیجه درخواست وام:	وام تایید شده یا نه (N/Y)