



# HomeWork

## Data Science

### DataScience\_Project

1. با استفاده از دیتابس loan موارد خواسته شده را پیاده سازی نمایید. کدهای خود را در یک فایل ipynb فایل با عنوان شماره تمرين و نام خودتان "YourName\_Project.zip" قرار داده و تاریخ ۴ خرداد ماه در پنل خود در سایت بارگذاری کنید (عدم ثبت نام شما در نام فایل می‌تواند باعث اشتباه در ثبت نمره و فیدبک شود).
2. در صورتی که سوال و یا ابهامی دارید در گروه چت تلگرامی بپرسید.

#### 1. پیش بینی واجدین شرایط وام

هدف: خودکار کردن فرآیند بررسی واجد شرایط وام بودن افراد بر اساس اطلاعات صاحبان حسابها. در این پروژه نیاز است ابتدا داده‌ها را پیش‌پردازش کنید، سپس مدل‌های مختلف را آموزش داده و در نهایت عملکرد آنها را ارزیابی نمایید.

#### بخش 1: پیش‌پردازش داده‌ها

(a) مدیریت مقادیر ناموجود: برخی ستون‌های دادگان این پروژه مقادیر ناموجود و گم شده دارند. در این بخش نیاز است با روش‌هایی که در این دوره یاد گرفته‌اید مشکل داده‌های 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 نرمال‌سازی داده‌های عددی در یادگیری مدل را توضیح دهید و مدل‌های بعدی را یک بار بدون نرمال‌سازی داده‌ها و یک بار با نرمال‌سازی 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 های پرپارامترها 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:

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