



پروژه اول درس مبانی یادگیری ماشین

مدل رگرسیون برای پیش‌بینی اعتبار مالی افراد

فایل credit_prediction.csv حاوی اطلاعات مختلف از مشتریان یک بانک است. اطلاعات ستون‌های این دیتا فریم به شرح زیر است:

Data Dictionary	
Description	Column
Client number. Unique identifier for the customer holding the account	CLIENTNUM
Demographic variable - Customer's Age in Years	Customer_Age
Demographic variable - M=Male, F=Female	Gender
Demographic variable - Number of dependents	Dependent_count
Demographic variable - Educational Qualification of the account holder	Education_Level
Demographic variable - Married, Single, Divorced, Unknown	Marital_Status
Demographic variable - Annual Income Category of the account holder	Income_Category
Product Variable - Type of Card (Blue, Silver, Gold, Platinum)	Card_Category
Period of relationship with bank	Months_on_book
Total no. of products held by the customer	Total_Relationship_count
No. of months inactive in the last 12 months	Months_Inactive_12_mon
No. of Contacts in the last 12 months	Contacts_Count_12_mon
Credit Limit on the Credit Card	Credit_Limit
Total Revolving Balance on the Credit Card	Total_Revolving_Bal
Change in Transaction Amount (Q4 over Q1)	Total_Amt_Chng_Q4_Q1
Total Transaction Amount (Last 12 months)	Total_Trans_Amt
Total Transaction Count (Last 12 months)	Total_Trans_Ct
Change in Transaction Count (Q4 over Q1)	Total_Ct_Chng_Q4_Q1
Average Card Utilization Ratio	Avg_Utilization_Ratio

شما باید ستون Credit_Limit را به عنوان ستون برچسب در نظر گرفته و یک مدل رگرسیونی مناسب پیشنهاد نمایید.

نکات:

- ارسال فایل متنی گزارش عملکرد و کد نهایی و همچنین مستندات کد، الگوریتم و مدل نهایی (به صورت متن یا ویدیو) الزامی است.
- حتما در مستندات خود روند الگوریتم نهایی حاوی پیش‌پردازش‌ها (حذف داده پرت، تمیز کردن دیتا، مهندسی و تحلیل ویژگی‌ها، مدل نهایی، تعیین ابرپارامترهای مدل نهایی و نحوه ارزیابی مدل و مقایسه با مدل‌های دیگر، چالش‌ها و پیشنهادات دیگر) را به طور کامل شرح دهید.
- فایل‌های خود را حداکثر تا تاریخ ۷ خردادماه به ایمیل drmmohamadkiani@gmail.com ارسال نمایید.
- به ازای هر روز تاخیر در ارسال، کسر نمره تا ۲۰ درصد منظور می‌گردد.
- به گزارش‌های مشابه نمره ای تعلق نمی‌گردد.
- محدودیتی در انتخاب پکیج و زبان برای حل مسئله وجود ندارد. اما در مستندات باید شرح داده شود.
- ارسال پروژه تک نفره می‌باشد و ارسال پروژه به صورت دو نفره تنها در صورتی مجاز است که مستندات و تحلیل‌ها بسیار جامع و کامل‌تر باشد وگرنه نمره پروژه تقسیم می‌گردد.

خوشدوین باشد. Drmkiani.ir