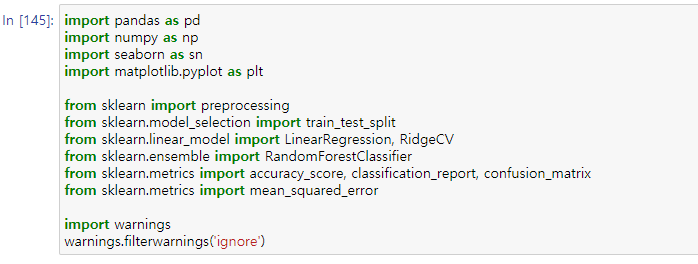
LOAN INTEREST RATE PREDICTION



Menggunakan beberapa package pengolahan data, matematika/kalkulasi, visualisasi data, sklearn(membantu pemrosesan data, skoring data dan memodelkan data) dan warnings (menghilangakan alert/output danger).

|  |  |  |
| --- | --- | --- |
| **Feature** | **Definition** | |
| Loan\_ID | A unique id for the loan. | |
| Loan\_Amount\_Requested | The listed amount of the loan applied for by the borrower. | |
| Length\_Employed | Employment length in years | |
| Home\_Owner | The home ownership status provided by the borrower during registration. Values are: Rent, Own, Mortgage, Other. | |
| Annual\_Income | The annual income provided by the borrower during registration. | |
| Income\_Verified | Indicates if income was verified, not verified, or if the income source was verified | |
| Purpose\_Of\_Loan | A category provided by the borrower for the loan request. | |
| Debt\_To\_Income | A ratio calculated using the borrower’s total monthly debt payments on the total debt obligations, excluding mortgage and the requested loan, divided by the borrower’s self-reported monthly income. | |
| Inquiries\_Last\_6Mo | The number of inquiries by creditors during the past 6 months. | |
| Months\_Since\_Deliquency | The number of months since the borrower's last delinquency. | |
| Number\_Open\_Accounts | The number of open credit lines in the borrower's credit file. | |
| Total\_Accounts | The total number of credit lines currently in the borrower's credit file | |
| Gender | Gender | |
| **Target** | **Definition** | |
| Interest\_Rate | Interest Rate category (1/2/3) of the loan application | |
| **Import Package**  **Dataset yang disediakan :**   1. Train.csv (Data Train) 2. Test.csv (Data Test)   **Data Exploration**    Dari informasi diatas kita bisa simpulkan tipe data masing-masing column dan ada beberapa column yang memiliki data yang bernilai null. Berikut column-column yang harus kita proses :   1. Length\_Employed 2. Home\_Owner 3. Annual\_Income 4. Months\_Since\_Deliquency   **Column yang memiliki data kategorikal :**        **Hasil visualisasi column-column yang memiliki data kategorikal**   |  |  | | --- | --- | | Visualisasi target variable : Interest\_Rate | Visualisasi Data Feature : Gender | |  |  | | Visualisasi Data Feature : Length\_Employed | Visualisasi Data Feature : Income\_Verified | |  |  | | Visualisasi Data Feature : Purpose\_Of\_Loan | Visualisasi Data Feature : Home\_Owner | |  |  |   **Data Preprocessing**   |  |  |  | | --- | --- | --- | | NO | FEATURE | NOTE | | 1 | Loan\_ID | Loan\_ID didrop karena tidak meberikan nilai prediksi terhadapa target, column ini akan kita gunakan saat nilai prediktif dari column Interest\_Rate telah dicapai. | | 2 | Loan\_Amount\_Requested | Konversi tipe data menjadi float | | 3 | Length\_Employed | Konversi value column berdasarkan kondisi tertentu dan mengubah ke bentuk tipe data float | | 4 | Home Owner | Mengisi data yang kosong sebagai “Own” karena yang paling mendekati missing value | | 5 | Annual\_Income | Secara langsung mengganti nilai yang kosong Annual\_Income dengan median/mean/modus, tetapi berasumsi bahwa Length\_Employed yang berbanding lurus dengan Annual\_Income, karena pendapatan untuk 10+ tahun karyawan tidak akan sebanding dengan karyawan 1 tahun (tentu saja ada outlier) sehingga kita lebih memilih median daripada mean | | 6 | Months\_Since\_Deliquency | Mengisi data yang kosong dengan median dari column Months\_Since\_Deliquency, karena data min-max data / beberapa data sangat jauh dari nilai mean. |   **Melakukan Model Selection**    Train Size = 70 ; Test Size = 30;  Random State = 42 ;  Stratify = Interest\_Rate  **Data Modeling** | | |

Dari beberapa percobaan parameter n\_estimator dan random state, maka pada saat n\_estimator = 100 dan random\_state = 42 memiliki accuracy score di **77%.**