**Zestaw danych do klasyfikacji zatwierdzania pożyczek**

Syntetyczne dane do klasyfikacji binarnej zatwierdzenia pożyczki

Zbiór danych zawiera 45 000 rekordów i 14 zmiennych, z których każda została opisana poniżej:

| **Column** | **Description** | **Type** |
| --- | --- | --- |
| person\_age | Age of the person | Float |
| person\_gender | Gender of the person | Categorical |
| person\_education | Highest education level | Categorical |
| person\_income | Annual income | Float |
| person\_emp\_exp | Years of employment experience | Integer |
| person\_home\_ownership | Home ownership status (e.g., rent, own, mortgage) | Categorical |
| loan\_amnt | Loan amount requested | Float |
| loan\_intent | Purpose of the loan | Categorical |
| loan\_int\_rate | Loan interest rate | Float |
| loan\_percent\_income | Loan amount as a percentage of annual income | Float |
| cb\_person\_cred\_hist\_length | Length of credit history in years | Float |
| credit\_score | Credit score of the person | Integer |
| previous\_loan\_defaults\_on\_file | Indicator of previous loan defaults | Categorical |
| loan\_status (target variable) | Loan approval status: 1 = approved; 0 = rejected | Integer |

Zbiór danych może być wykorzystywany do wielu celów:

Eksploracyjna analiza danych (EDA): Analiza kluczowych cech, wzorców dystrybucji i relacji w celu zrozumienia czynników ryzyka kredytowego.

Klasyfikacja: Tworzenie modeli predykcyjnych w celu sklasyfikowania zmiennej loan\_status (zatwierdzony/niezatwierdzony) dla potencjalnych wnioskodawców.

Regresja: Opracowanie modeli regresji w celu przewidywania zmiennej credit\_score w oparciu o atrybuty indywidualne i związane z pożyczką.

Należy pamiętać o kwestii danych z oryginalnych danych, takich jak instancja > 100 lat jako wiek.

Ten zestaw danych stanowi bogatą podstawę do zrozumienia czynników ryzyka finansowego i symulacji procesów modelowania predykcyjnego w celu zatwierdzenia pożyczki i oceny zdolności kredytowej.