Assignment 99 (Unit 99): Data Transformation

MSDS 422: Machine Learning
Author: Stefan Jenss
Instructor: Donald Wedding, PhD
Date: February 14th, 2023

Phase 1: Assignment Setup

Description of the numeric-only Data Frame:

Statistic	LOAN	MORTDUE	VALUE	YOJ	DEROG	DELINQ	CLAGE	NINQ	CLNO	DEBTINC
Count	5960.00	5442.00	5848.00	5445.00	5252.00	5380.00	5652.00	5450.00	5738.00	4693.00
Mean	18607.97	73760.82	101776.05	8.92	0.25	0.45	179.77	1.19	21.30	33.78
Std	11207.48	44457.61	57385.78	7.57	0.85	1.13	85.81	1.73	10.14	8.60
Min	1100.00	2063.00	8000.00	0.00	0.00	0.00	0.00	0.00	0.00	0.52
25%	11100.00	46276.00	66075.50	3.00	0.00	0.00	115.12	0.00	15.00	29.14
50%	16300.00	65019.00	89235.50	7.00	0.00	0.00	173.47	1.00	20.00	34.82
75%	23300.00	91488.00	119824.25	13.00	0.00	0.00	231.56	2.00	26.00	39.00
Max	89900.00	399550.00	855909.00	41.00	10.00	15.00	1168.23	17.00	71.00	203.31

Phase 2: Normalize the Data

2.1 Final `.head()` Results from Normalization Process:

	0	1	2	3	4
LOAN	1100	1300	1500	1500	1700
MORTDUE	25860.0	70053.0	13500.0	NaN	97800.0
VALUE	39025.0	68400.0	16700.0	NaN	112000.0
YOJ	10.5	7.0	4.0	NaN	3.0
DEROG	0.0	0.0	0.0	NaN	0.0
DELINQ	0.0	2.0	0.0	NaN	0.0
CLAGE	94.366667	121.833333	149.466667	NaN	93.333333
NINQ	1.0	0.0	1.0	NaN	0.0
CLNO	9.0	14.0	10.0	NaN	14.0
DEBTINC	NaN	NaN	NaN	NaN	NaN
nor_LOAN	0.0	0.002252	0.004505	0.004505	0.006757
nor_MORTDUE	0.059869	0.17105	0.028773	NaN	0.240856
nor_VALUE	0.03659	0.071234	0.010261	NaN	0.122655
nor_YOJ	0.256098	0.170732	0.097561	NaN	0.073171
nor_DEROG	0.0	0.0	0.0	NaN	0.0
nor_DELINQ	0.0	0.133333	0.0	NaN	0.0
nor_CLAGE	0.080777	0.104289	0.127942	NaN	0.079893
nor_NINQ	0.058824	0.0	0.058824	NaN	0.0
nor_CLNO	0.126761	0.197183	0.140845	NaN	0.197183
nor_DEBTINC	NaN	NaN	NaN	NaN	NaN
TARGET_BAD_FLAG	1	1	1	1	0
TARGET_LOSS_AMT	641.0	1109.0	767.0	1425.0	NaN
REASON	HomeImp	HomeImp	HomeImp	NaN	HomeImp
JOB	Other	Other	Other	NaN	Office

2.2 Calculating Normalized Values for `LOAN` Without Using MinMaxScalar:

LOAN	nor_LOAN	calc_Loan
1100	0.000000	0.000000
1300	0.002252	0.002252
1500	0.004505	0.004505
1500	0.004505	0.004505
1700	0.006757	0.006757

> The MinMaxScalar values and the calculated values are identical.

Phase 3: Standardize the Data

3.1 Final `.head()` Results from Standarization Process:

	0	1	2	3	4
LOAN	1100	1300	1500	1500	1700
MORTDUE	25860.0	70053.0	13500.0	NaN	97800.0
VALUE	39025.0	68400.0	16700.0	NaN	112000.0
YOJ	10.5	7.0	4.0	NaN	3.0
DEROG	0.0	0.0	0.0	NaN	0.0
DELINQ	0.0	2.0	0.0	NaN	0.0
CLAGE	94.366667	121.833333	149.466667	NaN	93.333333
NINQ	1.0	0.0	1.0	NaN	0.0
CLNO	9.0	14.0	10.0	NaN	14.0
DEBTINC	NaN	NaN	NaN	NaN	NaN
std_LOAN	-1.562299	-1.544453	-1.526606	-1.526606	-1.508759
std_MORTDUE	-1.077548	-0.083409	-1.355591	NaN	0.540771
std_VALUE	-1.093588	-0.581658	-1.482655	NaN	0.178177
std_YOJ	0.208329	-0.253822	-0.649951	NaN	-0.781994
std_DEROG	-0.300922	-0.300922	-0.300922	NaN	-0.300922
std_DELINQ	-0.398738	1.375631	-0.398738	NaN	-0.398738
std_CLAGE	-0.995304	-0.675189	-0.353132	NaN	-1.007348
std_NINQ	-0.107639	-0.686169	-0.107639	NaN	-0.686169
std_CLNO	-1.212866	-0.719675	-1.114228	NaN	-0.719675
std_DEBTINC	NaN	NaN	NaN	NaN	NaN
TARGET_BAD_FLAG	1	1	1	1	0
TARGET_LOSS_AMT	641.0	1109.0	767.0	1425.0	NaN
REASON	HomeImp	HomeImp	HomeImp	NaN	HomeImp
JOB	Other	Other	Other	NaN	Office

3.2 Manual Calculation of the StandardScaler for `std_LOAN`

LOAN	std_LOAN	calc_Loan
1100	-1.562299	-1.562168
1300	-1.544453	-1.544323
1500	-1.526606	-1.526478
1500	-1.526606	-1.526478
1700	-1.508759	-1.508633

> The calculated values are extremely similar to the StandardScaler numbers.