



An Application of AI to Credit Risk Modelling in the Auto Loan Industry



Presentation Highlights

- Business Context
- Understanding Credit Risk Modelling
- Credit Risk Modelling
- Data Preparation
- AI Technology : Logistic Regression
- Predicting Auto Loan Defaults
- Conclusions

Business Context

- An auto financing company that has always relied on credit risk expert in predicting loan defaults will now like to leverage on AI technology to better and quickly predict loan defaults as it plans to massively increase its market share.
- It will also like to know the main determinant of defaults in order to better manage its risks.
- A database of customers' loans containing loan information, loanee information and credit bureau data and history is provided.

Understanding Credit Risk Modelling

Some Terminologies:

- Credit Risk is the possibility or risk that the borrower will not fully repay the loan. Calculated as the difference between the risk of lending a customer money and the risk of buying a government bond
- The likelihood that a customer will default on a loan is the probability of default (PD)
- The expected loss from a loan is the $PD * Loss\ Severity * Exposure\ at\ Default$

Credit Risk Modelling

- The loss severity which is calculated as $1 - \text{recovery rate}$ is determined by the how much the company is able to recovery if a customer defaults and this will be dependent on a host of factors.
- We will focus on using AI to predict the probability of defaults given the data available

Data Preparation

- 40 features such as Date.of.Birth, asset_cost, and State_ID
- 233 154 rows
- Convert the date column to datetime data type
- Fill up missing values
- Divide into test and train datasets
- Drop unnecessary columns

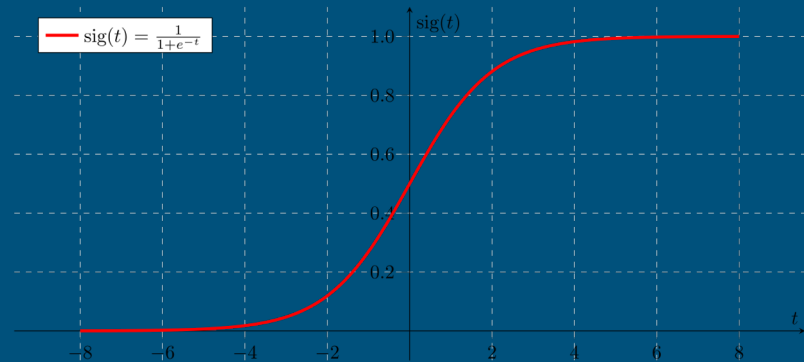
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 186523 entries, 50119 to 71787
Data columns (total 40 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   UniqueID                             186523 non-null int64
1   disbursed_amount                     186523 non-null int64
2   asset_cost                           186523 non-null int64
3   ltv                                  186523 non-null float64
4   branch_id                           186523 non-null int64
5   supplier_id                         186523 non-null int64
6   manufacturer_id                     186523 non-null int64
7   Current_pincode_ID                  186523 non-null int64
8   Date.of.Birth                       186523 non-null int64
9   Employment.Type                     180349 non-null object
10  DisbursalDate                       186523 non-null object
11  State_ID                            186523 non-null int64
12  Employee_code_ID                    186523 non-null int64
13  MobileNo_Avl_Flag                  186523 non-null int64
14  Aadhar_flag                         186523 non-null int64
15  PAN_flag                           186523 non-null int64
16  VoterID_flag                       186523 non-null int64
17  Driving_flag                       186523 non-null int64
18  Passport_flag                      186523 non-null int64
19  PERFORM_CNS.SCORE                   186523 non-null int64
20  PERFORM_CNS.SCORE.DESCRPTION       186523 non-null object
21  PRI.NO.OF.ACCTS                     186523 non-null int64
22  PRI.ACTIVE.ACCTS                    186523 non-null int64
23  PRI.OVERDUE.ACCTS                   186523 non-null int64
24  PRI.CURRENT.BALANCE                 186523 non-null int64
25  PRI.SANCTIONED.AMOUNT               186523 non-null int64
```

AI Technology: Logistic Regression

- As our target features is dichotomous (default/non default) and we are also interested in knowing the probability of default predicted by our model, we chose to apply a logistic regression model to our use case.
- The prediction of the probability of defaults can be used in calculating expected loss from our loan receivable (a reporting requirement from regulators) as well as in determining if a customer will default on its auto loan on the due date.

AI Technology: Logistic Regression

- Regression analysis to conduct when the dependent variable is dichotomous (binary)
- Describe data and to explain the relationship between one dependent binary variable and one or more independent variables
- Developed and popularized primarily by Joseph Berkson, beginning in Berkson (1944)



Prediction of Auto Loan Defaults

Our Model

- The `predict_proba()` method of our logistic regression is able to predict probability of defaults which can be used for expected loss computation and loan status reclassification
- To determine an appropriate threshold for loan default status classification we consider that It's important for our model to be able to predict true loan defaults than true non default loans as risk management actions will be based on default predictions
- We determine the optimal loan default threshold (0.211915) using the largest J statistics

Prediction of Auto Loan Defaults

Validation

The recall for defaults is 0.66 meaning 66% of our actual loan defaults were predicted correctly.

	precision	recall	f1-score	support
Non-Default	0.84	0.49	0.62	36585
Default	0.26	0.66	0.37	10046
accuracy			0.52	46631
macro avg	0.55	0.57	0.49	46631
weighted avg	0.71	0.52	0.56	46631

Prediction of Auto Loan Defaults

- Estimated default probability of some entries:

No Default Probability	Default Probability
0.7054146	0.2945854
0.81272068	0.18727932

- Confusion matrix

17817	18768
3451	0

Conclusion

The model is able to identify loans that will be in default on the loans due date, hence Credit risk department can target risk management actions on this loans thereby reducing the expected loss from this loans that would have adverse impacted our profits.