

Credit Risk Analysis

Background – Credit Risk Analysis

CredX gives credit cards to thousands of people every year, of which **approx. 4% default**. The defaulters form the largest fraction of the portfolio's loss (credit loss).

The objectives of the analysis are to:

- Identify the most important variables affecting likelihood of default
- Build an application scorecard to identify the likely defaulters at the application stage using predictive models
- Estimate the potential financial benefits of using the models for auto-approval of credit cards

Credit Risk Analysis - Flow of Topics

The analysis is divided into 5 parts:

- **Data Understanding** – Demographic and Credit bureau information
- **Identifying important variables using** Exploratory Data Analysis
- **Predictive modelling**
 - Modelling on demographic data only
 - Modelling on combined data of demographic and credit bureau variables
- **Application scorecard**
 - Identifying the optimal score for rejecting the applicant
- **Financial Benefits**
 - Assessing the potential benefits of using predictive models for auto-approval

-
- **Data Understanding**
 - Identifying important variables
 - Predictive modelling
 - Application scorecard
 - Financial Benefits

Data Understanding – Demographic and Credit Bureau Data

<div><div>Demographic Data</div><div>Provided by applicants at the time of credit card application.</div></div>	Application Information*
	Age
	Income
	Gender
	Marital Status
	Education
<div><div>Credit Bureau Data</div><div>Provided by credit bureau agency of every individual. The data contains Information related to applicants' previous loans, credit cards etc.</div></div>	Credit Bureau Information**
	Outstanding balance
	Past due 30,60,90 DPD
	Total trades
	Number of inquiries
	Presence of home loan

* Demographic Data contains 12 attributes. Only few are shown in the table

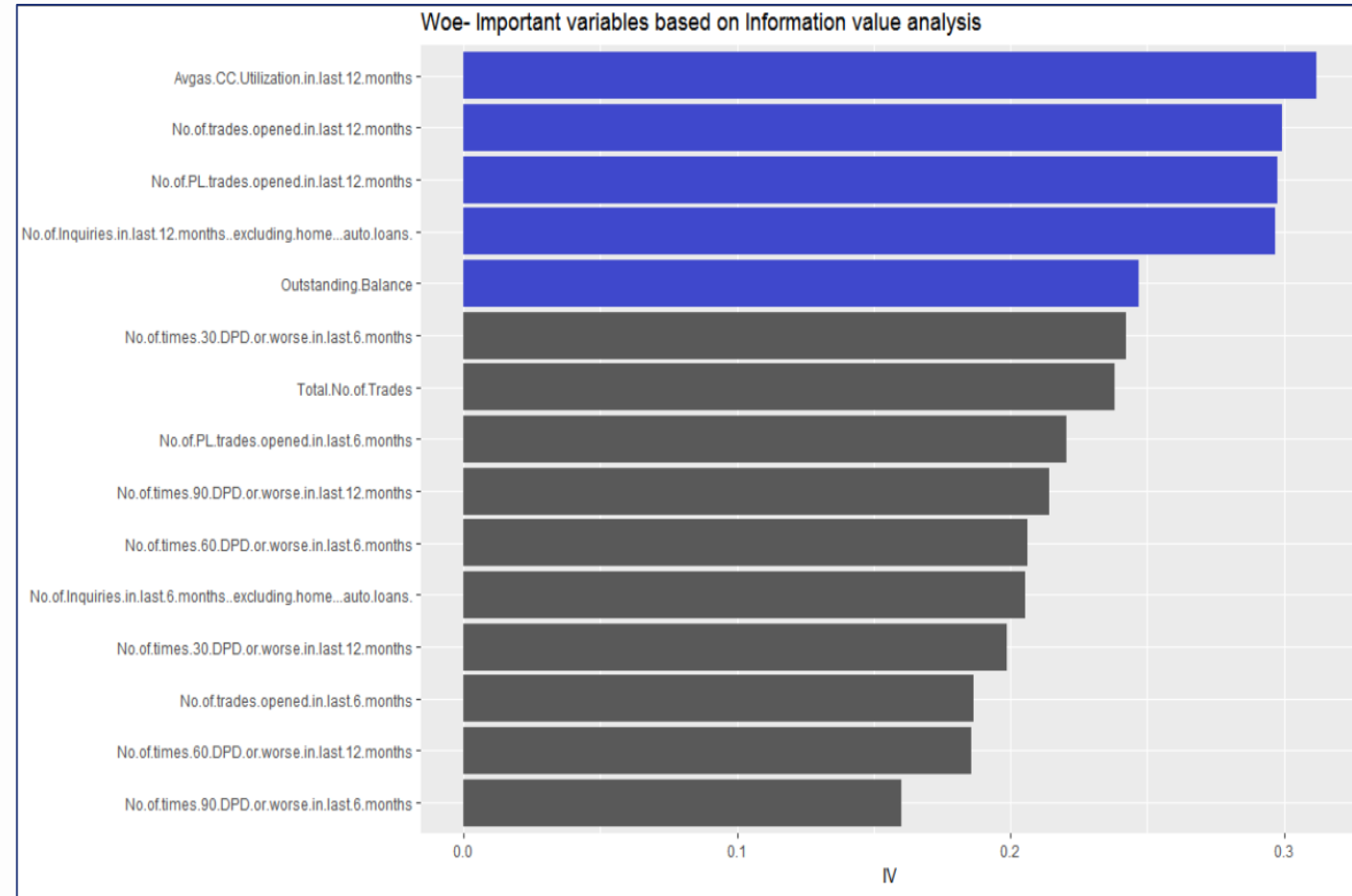
** Credit Bureau Data contains 19 attributes. Only few are shown in the table

-
- Data Understanding
 - **Identifying important variables**
 - Predictive modelling
 - Application scorecard
 - Financial Benefits

Identifying Important Variables: Average credit utilisation, Trades opened, Inquiries and Outstanding Balance

The most crucial variables seem to be:

- Average credit utilisation in last 12 months
- Number of trades opened in last 12 months
- Number of PL(personal loan) trades opened in last 12 months
- Number of Inquiries in last 12 months (excluding home auto loans)
- Outstanding balance



plot only contains variables with “Strong” and “Medium” IV(Information value)

-
- Data Understanding
 - Identifying important variables
 - **Predictive modelling**
 - Application scorecard
 - Financial Benefits

Predictive Modelling – Best Model: Random Forest*: Accuracy: 72% , Sensitivity: 75% and Specificity: 72%

- Model identifies 75% of defaulters correctly
- Captures 80% defaulters in top 4 deciles

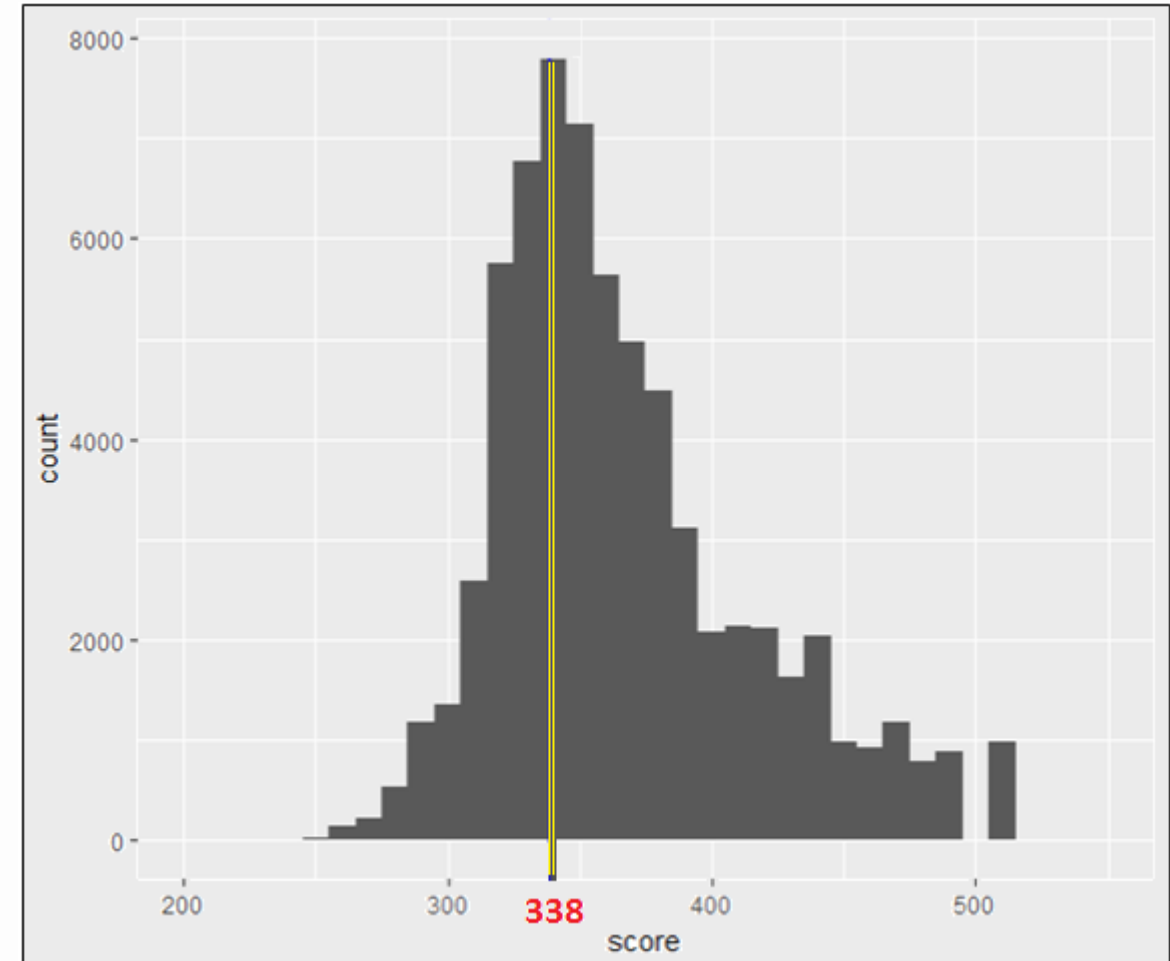
bucket	total	Total Bad	Cum- Bad	Gain	Lift
1	6951	1739	1739	59.2	5.9
2	6950	272	2011	68.4	3.4
3	6950	195	2206	75.1	2.5
4	6950	169	2375	80.8	2.0
5	6950	155	2530	86.1	1.7
6	6950	122	2652	90.3	1.5
7	6950	95	2747	93.5	1.3
8	6950	85	2832	96.4	1.2
9	6950	58	2890	98.4	1.1
10	6950	48	2938	100.0	1.0

*Random Forest model trained on balanced data

-
- Data Understanding
 - Identifying important variables
 - Predictive modelling
 - **Application scorecard**
 - Financial Benefits

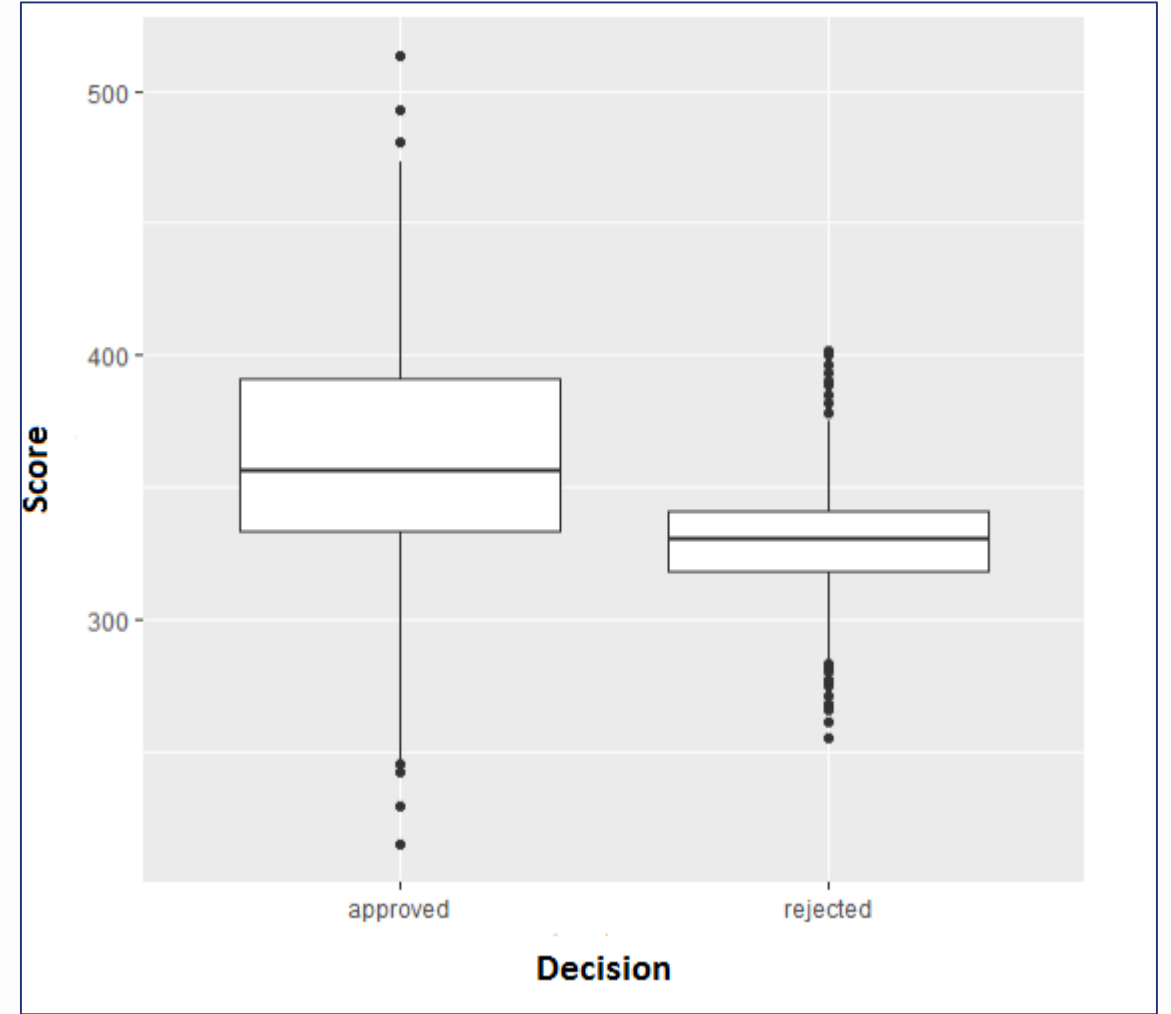
Application Scorecard (master population): Score varies between 200 to 530; Cut-off score - 338

- Cut-off: 338 is the baseline for providing credit card to the customers



Application Scorecard (rejected population) : 70% of defaulters correctly identified

- Average score of rejected population is less than the average score of approved* population
- Total rejected applications by bank: **1423**
- Identified correctly at cut-off score by model: **1006**



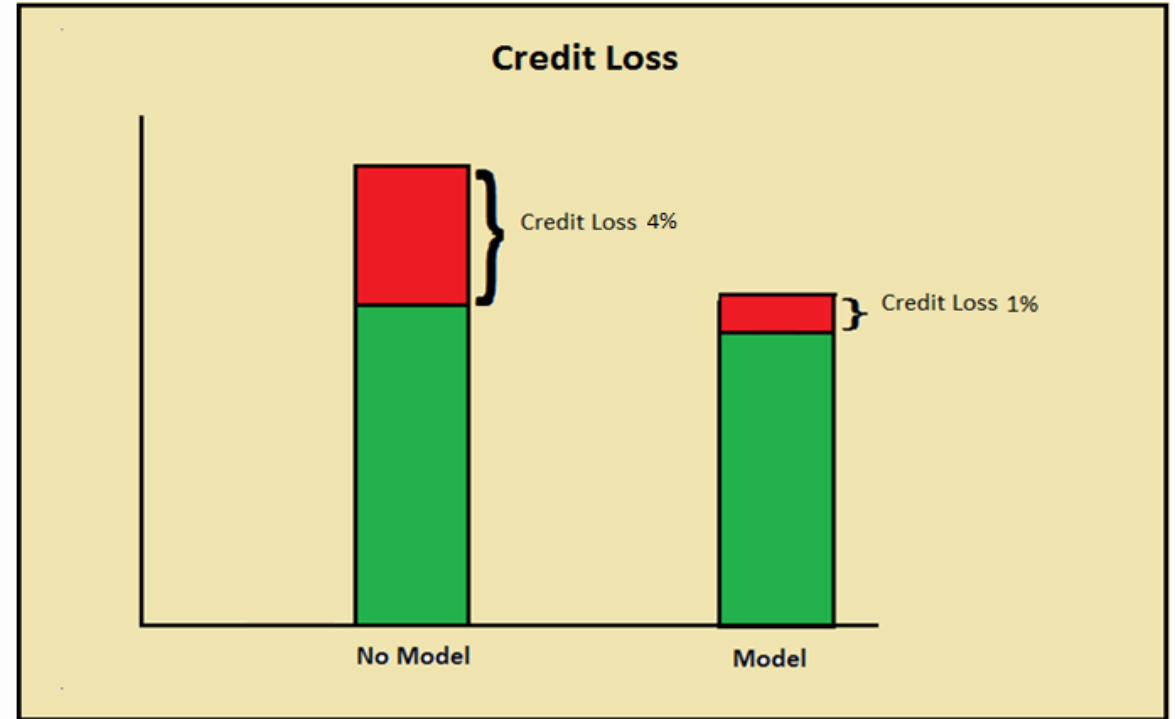
*Approved population (master data) is a population for which the application is accepted by bank

-
- Data Understanding
 - Identifying important variables
 - Predictive modelling
 - Application scorecard
 - **Financial Benefits**

Credit Loss*: Reduced credit loss from 4% customers to 1% customers

- Credit loss no model = 4%
- Credit loss with model = 1%

Credit Loss Saved: 3%



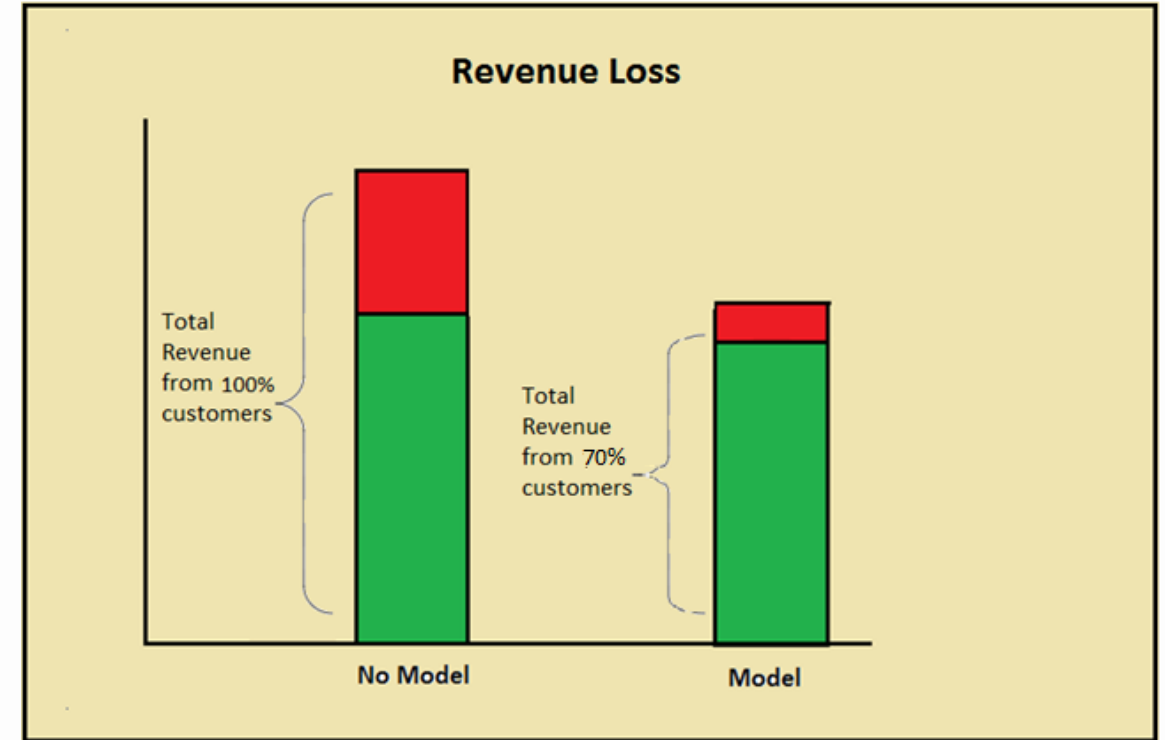
Confusion Matrix		Actual Defaults	
		Good Customers(0)	Bad Customers(1)
Predicted Defaults	Good Customers(0)	47938	732
	Bad Customers (1)	18625	2206

* The loss occurred from the bad customers

Revenue Loss*: Reducing 30% revenue (Auto-approval)

- Revenue no model = **100%**
- Revenue with model = **70%**

Revenue Loss : 30%



Confusion Matrix		Actual Defaults	
		Good Customers(0)	Bad Customers(1)
Predicted Defaults	Good Customers(0)	47938	732
	Bad Customers (1)	18625	2206

* The revenue loss is occurred by wrongly identified “bad” to the good customers