



SocBiz IITR Analytics

Credit Card Default Strategy &
Prediction

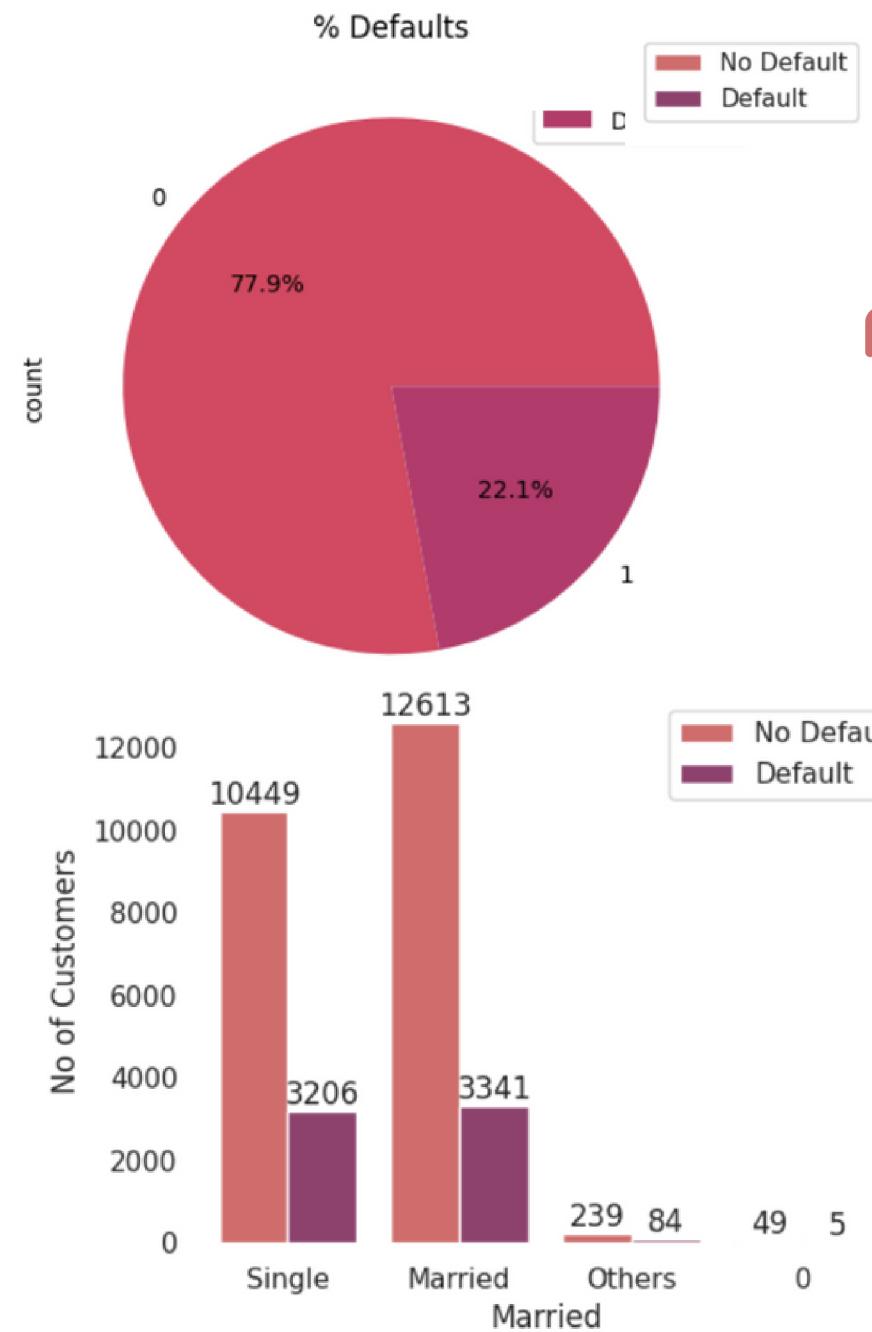
Agam Pandey,
Member- Analytics

Executive Summary

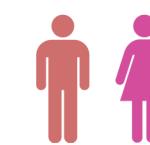
- Credit card defaults are affected by factors including the Gender of the customer, **Male customers** between the **ages of 23-30yr tend to default more** on bill payments compared to 38% of Female CC holders.
- 25% Married and 20% Single customers are prone to CC default.
- **Limit of Credit card and % usage by cardholders impact the default.** Customers with less than 15% card limit usage default more.
- **Defaulters are likely to spend less money compared to others**
- Customer % spending, minimum bill payment, and education background have a resounding effect on CC defaults
- **SelectKBest and Correlation Matrix** were used to **find features impacting CC Default** the most.
- 16/24 features gave higher scores by ANOVA statistical test, thus helping us identify important features for Default prediction.
- **Random Forest Classifier model** was used to predict Default, the model had higher **Accuracy(82%), Recall (0.78) and Precision (0.80)** for 0/1 classes compared to Logistic Regression, Naive Bayes, Decision Tree and XGBoost models
- Interactive **Web Application using StreamLit and ngrok** (to host the app were used).

CC Defaults are affected by Gender, Age, Marital Status and Card Limit

Predict Credit Card Defaulters and Identify key features impacting the Defaults thereby helping in decision-making process in the organisation



22.1% Customers defaulted on CC payments during May 2005-September 2005

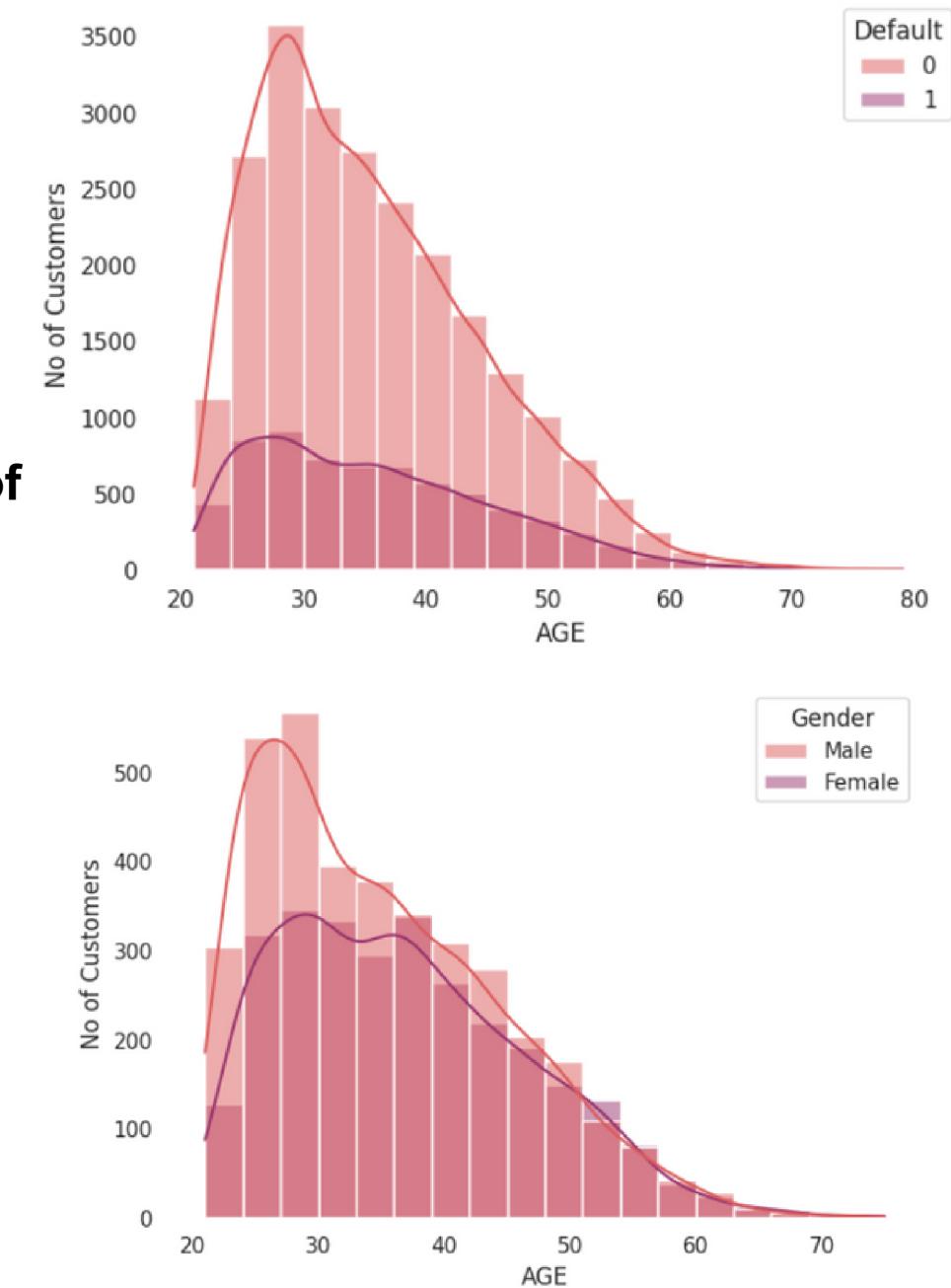


Defaulters by Gender shows that **31% Male** and **15.8% Female** CC holder fail to pay bills.

The analysis of Marital status shows that **25% of married** and **21% of Single** customers tend to **Default on CCs**

Diving deeper, Age is an important factor for CC Defaults customers of **Age group 23-30yrs are more likely to default** than others.

Further, in the **age group of 23-30, 62% Males and 38% Females default**



Analysis and insights done from 30k row Dataset-

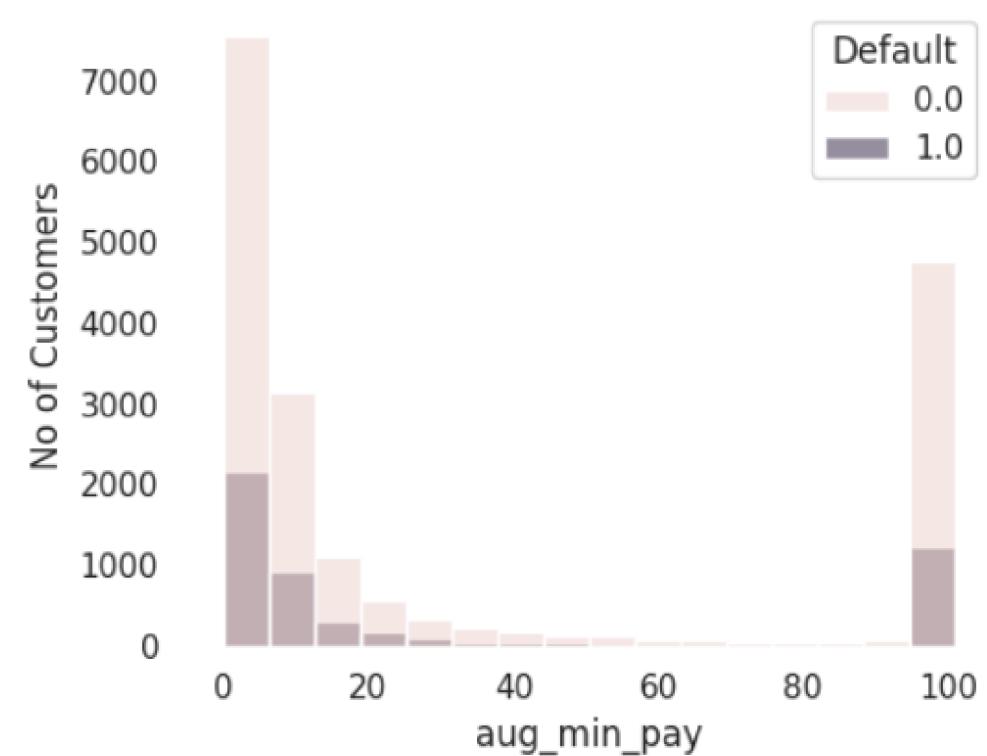
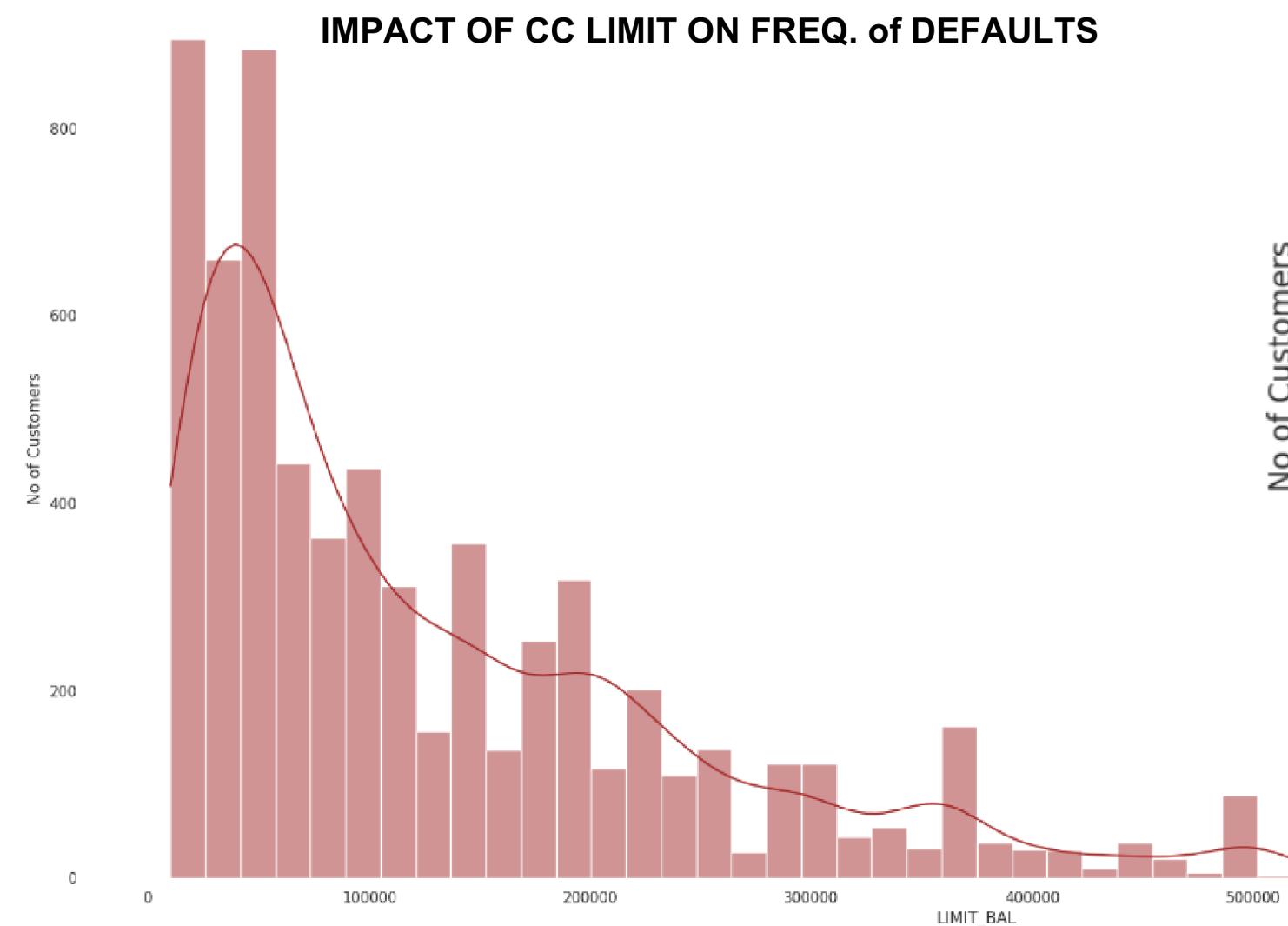
https://iitracin-my.sharepoint.com/:x/g/personal/h_bidasaria_hs_iitr_ac_in/EduGV6I4hNRGs6ITnlefF1QBO5AXdeSxCFZxUWErPEplvg?e=lWMtRe

Customer % spending, minimum bill payment, and education background have resounding effects on CC defaults

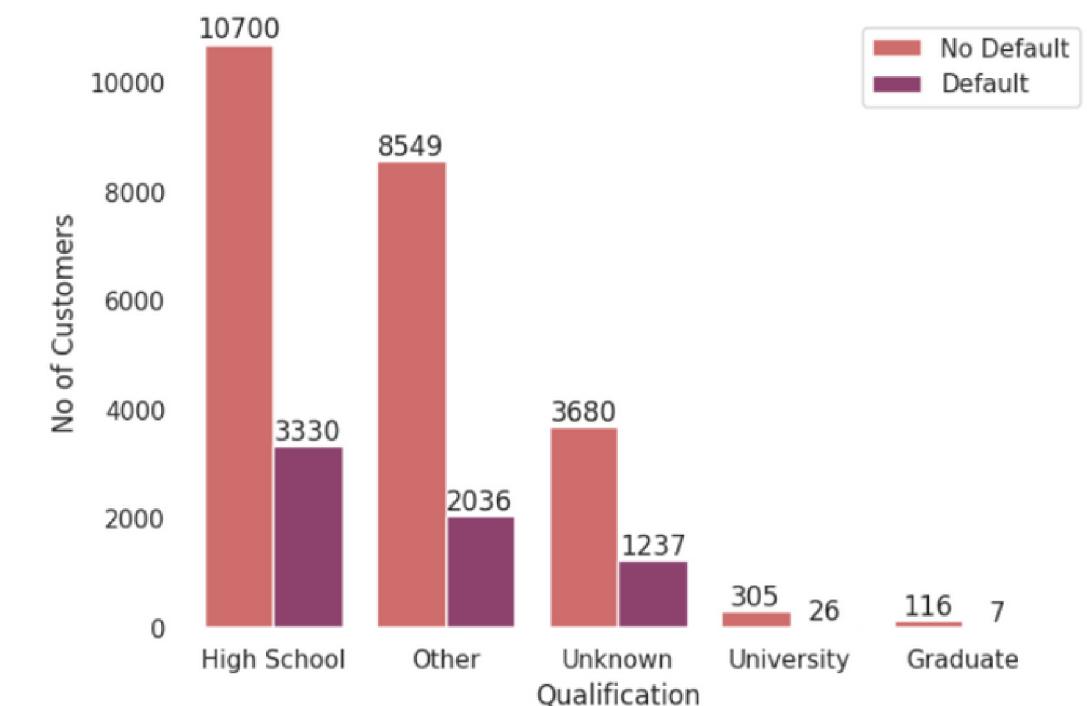
Other factors impacting Defaults are the **CC limits**, Credit Card limit of **below 100k negatively impacts bill payments** by customers.

Analyzing deeper into the CC limit, we find that **customers of age group 23-30 with CC limit < 100k default more than others**.

Defaulters tend to spend less than non Defaulters on CC with a **% usage of less than 20%** of card's limit



Customers who paid a minimum of only 15% in August went on to default their payment by September



Based on Qualification, **High school** passed and customers with “**Unknown**” and “**Other**” categories **default** on payments

SelectKBest and Correlation Matrix were used to find features impacting CC Default the most

Features and Factors impacting the most (mathematically explained)

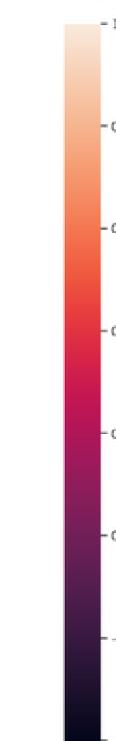
2 statistical methodology followed to find impacting features on Default (0 or 1)

	Feature	Scores
6	PAYO	2436.326083
7	PAY2	1993.725674
2	Gender	1613.810845
4	Married	1487.281985
8	PAY3	1453.173021
9	PAY4	1231.835574
1	LIMIT_BAL	1193.418191
10	PAY5	1068.191732
11	PAY6	841.438727
18	Bill1	402.036314
3	Qualification	372.155027
20	Bill3	258.338059
22	Bill5	258.019308
19	Bill2	236.775923
23	Bill6	236.369224
21	Bill4	228.267744
12	Amount1	15.491035
0	ID	12.748847
13	Amount2	5.610944
14	Amount3	4.228073
15	Amount4	1.917876
16	Amount5	0.444097
17	Amount6	0.055904
5	AGE	0.030146

ID	LIMIT_BAL	Gender	Education	Married	AGE	PAYO	PAY2	PAY3	PAY4	PAY5	PAY6	Amount1	Amount2	Amount3	Amount4	Amount5	Amount6	Bill1	Bill2	Bill3	Bill4	Bill5	Bill6	Default					
						0.026	0.019	0.039	-0.029	0.019	0.027	0.018	0.013	0.029	0.014	0.018	0.019	0.018	0.024	0.04	0.017	0.017	0.0098	0.0085	0.039	0.00790	0.00066	0.003	-0.014

Above analysis shows that for SelectKBest Features impacting the most on Defaults are --

- Repayment status in current and past months ,
- Gender,
- Marital Status,
- Qualification and,
- Billing Amount per month on CC



ML Model for Prediction of Default using Binary Classifiers

Logistic Regression

Random Forest Classifier

Naive Bayes

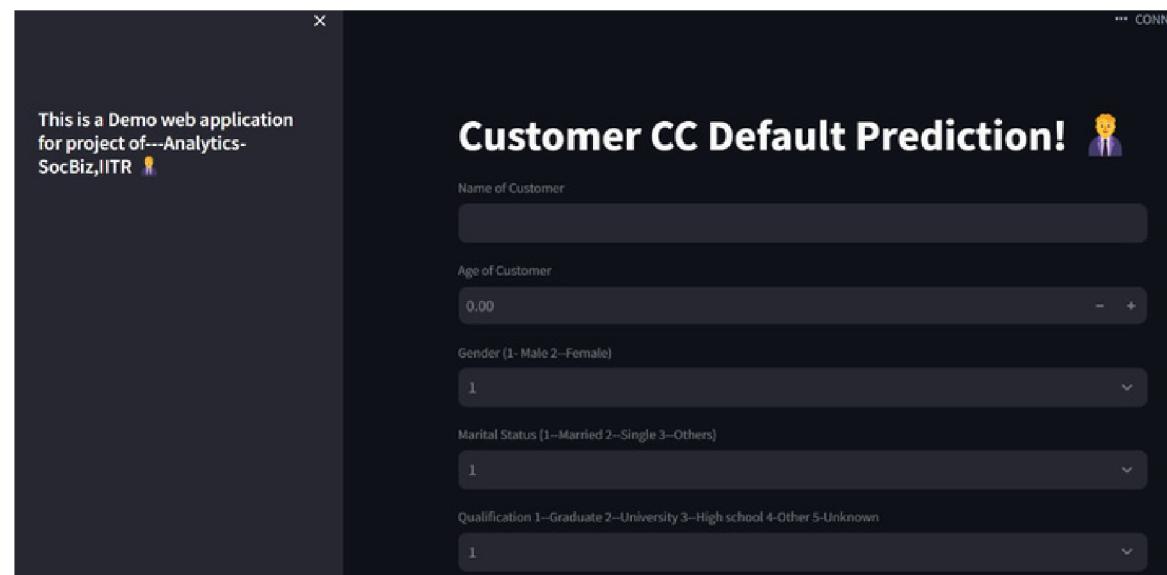
XGBoost

Decision Tree Classifier

The above models were used to predict Default of CC based on input features of 16 best impacting feature selected by SelectKBest

Random Forest Classifier model to predict Default and Web Application using StreamLit

	Logistic Regression	Random Forest Classifier	Naive Bayes	XGBoost	Decision Tree Classifier																																																																																																																																																																																														
Model Accuracy	65.3%	81.9%	54.5%	50.2%	74.07%																																																																																																																																																																																														
Confusion Matrix	[4638, 2406] [2452, 4514]	[6051, 993] [1529, 5437]	[1052 5992] [372 6594]	[7044 0] [6966 0]	[5163 1881] [1751 5215]																																																																																																																																																																																														
Classification Report	<table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.65</td><td>0.66</td><td>0.66</td><td>7044</td></tr><tr><td>1</td><td>0.65</td><td>0.65</td><td>0.65</td><td>6966</td></tr></tbody></table> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.80</td><td>0.86</td><td>0.83</td><td>7044</td></tr><tr><td>1</td><td>0.85</td><td>0.78</td><td>0.81</td><td>6966</td></tr></tbody></table> <table><thead><tr><th></th><th>accuracy</th><th>macro avg</th><th>weighted avg</th><th></th></tr></thead><tbody><tr><td>accuracy</td><td>0.65</td><td>0.65</td><td>0.65</td><td>14010</td></tr><tr><td>macro avg</td><td>0.65</td><td>0.65</td><td>0.65</td><td>14010</td></tr><tr><td>weighted avg</td><td>0.65</td><td>0.65</td><td>0.65</td><td>14010</td></tr></tbody></table>		precision	recall	f1-score	support	0	0.65	0.66	0.66	7044	1	0.65	0.65	0.65	6966		precision	recall	f1-score	support	0	0.80	0.86	0.83	7044	1	0.85	0.78	0.81	6966		accuracy	macro avg	weighted avg		accuracy	0.65	0.65	0.65	14010	macro avg	0.65	0.65	0.65	14010	weighted avg	0.65	0.65	0.65	14010	<table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.80</td><td>0.86</td><td>0.83</td><td>7044</td></tr><tr><td>1</td><td>0.85</td><td>0.78</td><td>0.81</td><td>6966</td></tr></tbody></table> <table><thead><tr><th></th><th>accuracy</th><th>macro avg</th><th>weighted avg</th><th></th></tr></thead><tbody><tr><td>accuracy</td><td>0.82</td><td>0.82</td><td>0.82</td><td>14010</td></tr><tr><td>macro avg</td><td>0.82</td><td>0.82</td><td>0.82</td><td>14010</td></tr><tr><td>weighted avg</td><td>0.82</td><td>0.82</td><td>0.82</td><td>14010</td></tr></tbody></table>		precision	recall	f1-score	support	0	0.80	0.86	0.83	7044	1	0.85	0.78	0.81	6966		accuracy	macro avg	weighted avg		accuracy	0.82	0.82	0.82	14010	macro avg	0.82	0.82	0.82	14010	weighted avg	0.82	0.82	0.82	14010	<table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.74</td><td>0.15</td><td>0.25</td><td>7044</td></tr><tr><td>1</td><td>0.52</td><td>0.95</td><td>0.67</td><td>6966</td></tr></tbody></table> <table><thead><tr><th></th><th>accuracy</th><th>macro avg</th><th>weighted avg</th><th></th></tr></thead><tbody><tr><td>accuracy</td><td>0.55</td><td>0.55</td><td>0.55</td><td>14010</td></tr><tr><td>macro avg</td><td>0.63</td><td>0.55</td><td>0.46</td><td>14010</td></tr><tr><td>weighted avg</td><td>0.63</td><td>0.55</td><td>0.46</td><td>14010</td></tr></tbody></table>		precision	recall	f1-score	support	0	0.74	0.15	0.25	7044	1	0.52	0.95	0.67	6966		accuracy	macro avg	weighted avg		accuracy	0.55	0.55	0.55	14010	macro avg	0.63	0.55	0.46	14010	weighted avg	0.63	0.55	0.46	14010	<table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.50</td><td>1.00</td><td>0.67</td><td>70</td></tr><tr><td>1</td><td>0.00</td><td>0.00</td><td>0.00</td><td>69</td></tr></tbody></table> <table><thead><tr><th></th><th>accuracy</th><th>macro avg</th><th>weighted avg</th><th></th></tr></thead><tbody><tr><td>accuracy</td><td>0.50</td><td>0.50</td><td>0.50</td><td>140</td></tr><tr><td>macro avg</td><td>0.25</td><td>0.50</td><td>0.33</td><td>140</td></tr><tr><td>weighted avg</td><td>0.25</td><td>0.50</td><td>0.34</td><td>140</td></tr></tbody></table>		precision	recall	f1-score	support	0	0.50	1.00	0.67	70	1	0.00	0.00	0.00	69		accuracy	macro avg	weighted avg		accuracy	0.50	0.50	0.50	140	macro avg	0.25	0.50	0.33	140	weighted avg	0.25	0.50	0.34	140	<table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>0</td><td>0.75</td><td>0.73</td><td>0.74</td><td>7044</td></tr><tr><td>1</td><td>0.73</td><td>0.75</td><td>0.74</td><td>6966</td></tr></tbody></table> <table><thead><tr><th></th><th>accuracy</th><th>macro avg</th><th>weighted avg</th><th></th></tr></thead><tbody><tr><td>accuracy</td><td>0.74</td><td>0.74</td><td>0.74</td><td>14010</td></tr><tr><td>macro avg</td><td>0.74</td><td>0.74</td><td>0.74</td><td>14010</td></tr><tr><td>weighted avg</td><td>0.74</td><td>0.74</td><td>0.74</td><td>14010</td></tr></tbody></table>		precision	recall	f1-score	support	0	0.75	0.73	0.74	7044	1	0.73	0.75	0.74	6966		accuracy	macro avg	weighted avg		accuracy	0.74	0.74	0.74	14010	macro avg	0.74	0.74	0.74	14010	weighted avg	0.74	0.74	0.74	14010
	precision	recall	f1-score	support																																																																																																																																																																																															
0	0.65	0.66	0.66	7044																																																																																																																																																																																															
1	0.65	0.65	0.65	6966																																																																																																																																																																																															
	precision	recall	f1-score	support																																																																																																																																																																																															
0	0.80	0.86	0.83	7044																																																																																																																																																																																															
1	0.85	0.78	0.81	6966																																																																																																																																																																																															
	accuracy	macro avg	weighted avg																																																																																																																																																																																																
accuracy	0.65	0.65	0.65	14010																																																																																																																																																																																															
macro avg	0.65	0.65	0.65	14010																																																																																																																																																																																															
weighted avg	0.65	0.65	0.65	14010																																																																																																																																																																																															
	precision	recall	f1-score	support																																																																																																																																																																																															
0	0.80	0.86	0.83	7044																																																																																																																																																																																															
1	0.85	0.78	0.81	6966																																																																																																																																																																																															
	accuracy	macro avg	weighted avg																																																																																																																																																																																																
accuracy	0.82	0.82	0.82	14010																																																																																																																																																																																															
macro avg	0.82	0.82	0.82	14010																																																																																																																																																																																															
weighted avg	0.82	0.82	0.82	14010																																																																																																																																																																																															
	precision	recall	f1-score	support																																																																																																																																																																																															
0	0.74	0.15	0.25	7044																																																																																																																																																																																															
1	0.52	0.95	0.67	6966																																																																																																																																																																																															
	accuracy	macro avg	weighted avg																																																																																																																																																																																																
accuracy	0.55	0.55	0.55	14010																																																																																																																																																																																															
macro avg	0.63	0.55	0.46	14010																																																																																																																																																																																															
weighted avg	0.63	0.55	0.46	14010																																																																																																																																																																																															
	precision	recall	f1-score	support																																																																																																																																																																																															
0	0.50	1.00	0.67	70																																																																																																																																																																																															
1	0.00	0.00	0.00	69																																																																																																																																																																																															
	accuracy	macro avg	weighted avg																																																																																																																																																																																																
accuracy	0.50	0.50	0.50	140																																																																																																																																																																																															
macro avg	0.25	0.50	0.33	140																																																																																																																																																																																															
weighted avg	0.25	0.50	0.34	140																																																																																																																																																																																															
	precision	recall	f1-score	support																																																																																																																																																																																															
0	0.75	0.73	0.74	7044																																																																																																																																																																																															
1	0.73	0.75	0.74	6966																																																																																																																																																																																															
	accuracy	macro avg	weighted avg																																																																																																																																																																																																
accuracy	0.74	0.74	0.74	14010																																																																																																																																																																																															
macro avg	0.74	0.74	0.74	14010																																																																																																																																																																																															
weighted avg	0.74	0.74	0.74	14010																																																																																																																																																																																															



Random Forest Classifier performs the best among other binary classification algorithms with **accuracy of 82%**

The confusion matrix shows that RFC model has a **Precision of 0.80** i.e, it classifies 80/100 prediction of Default were correctly classified, **Recall of 0.78** shows that 78% correct prediction of Default were made of the actual Default

An interactive **Web application** made using **StreamLit** that uses **Random Forest Classifier** to predict **Default** of a customer based on the inputs like Gender, CC Limit, Age,etc.



SocBiz IITR Analytics

THANK YOU!

ipynb- <https://www.kaggle.com/agampy/socbiz-analproj>