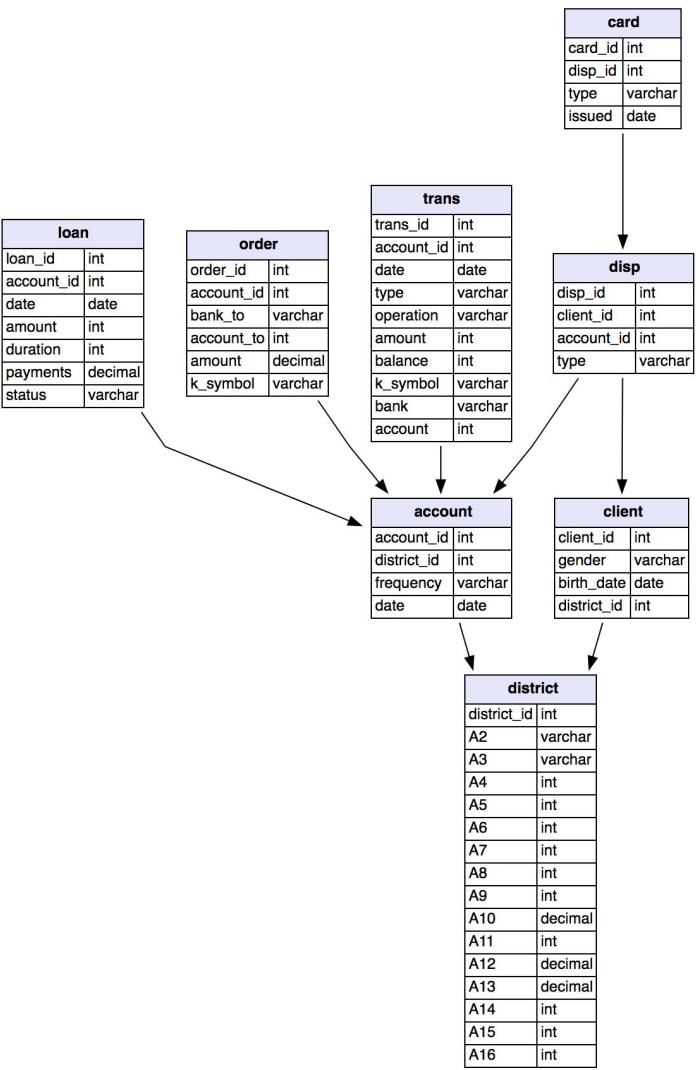


# Bank Loan Default Prediction

Taehoon Roh

# Data



# Mysql + Ssqlalchemy + Pandas + Numpy



loan_id	account_id	date	amount	duration	payments	status	frequency	district_id	A4	...	gender	last_balance	min_balance	mean_balan	std_balance	debited_amount	type	issued_date	age	card_type
0	4959	2 1994-01-05	80952	24	3373	A	POPLATEK MESICNE	1	1204953	...	Men	20575	1100	32372.2909	12058.935712	10639	no card	no card	48	no_card
1	4961	19 1996-04-29	30276	12	2523	B	POPLATEK MESICNE	21	103347	...	Women	15854	715	25197.1375	15039.321182	2523	no card	no card	57	no_card
2	4962	25 1997-12-08	30276	12	2523	A	POPLATEK MESICNE	68	228848	...	Men	31458	900	62800.2969	22541.820034	10614	no card	no card	35	no_card
3	4967	37 1998-10-14	318480	60	5308	D	POPLATEK MESICNE	20	70646	...	Men	16368	-1011	39954.0344	23451.696220	10306	no card	no card	46	no_card
4	4968	38 1998-04-19	110736	48	2307	C	POPLATEK TYDNE	19	51428	...	Women	37417	13841	31383.5818	10950.723180	9667	no card	no card	58	no_card
5	4973	67 1996-05-02	165960	24	6915	A	POPLATEK MESICNE	16	93931	...	Women	23704	700	52523.3360	20955.670441	13753	no card	no card	51	no_card
6	4986	97 1997-08-10	102876	12	8573	A	POPLATEK MESICNE	74	323870	...	Men	39080	900	33108.0672	12629.232624	12438	classic	1998-06-23	55	no_card

682 rows × 34 columns

# Target Variable

682 loans

A : contract finished, no problems,

B : contract finished, loan not paid,

C : running contract, OK so far,

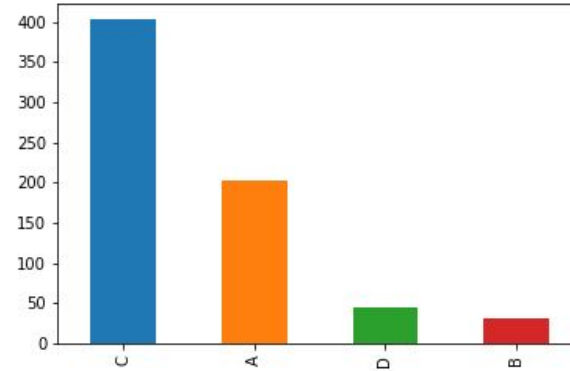
D : running contract, client in debt

0 : A + C

1 : B + D

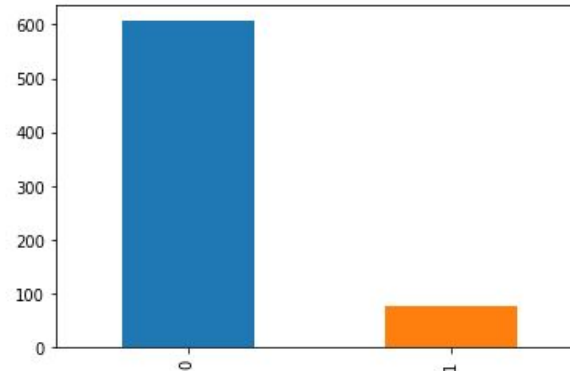
Imbalanced -> Smote

Status



A : 203  
B : 31  
C : 403  
D : 45

Default

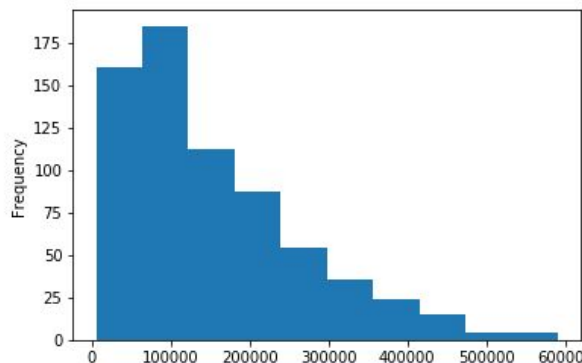


0 : 606  
1 : 76

# Features

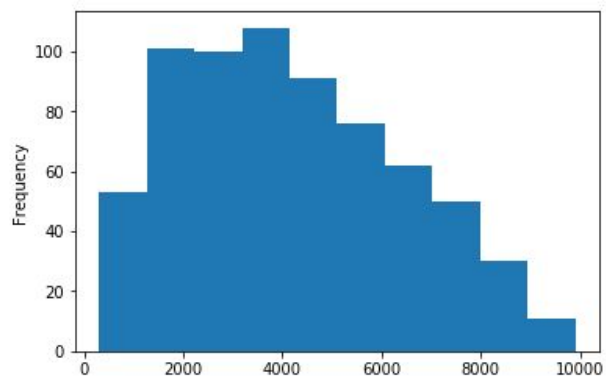
## Loan

Loan amount



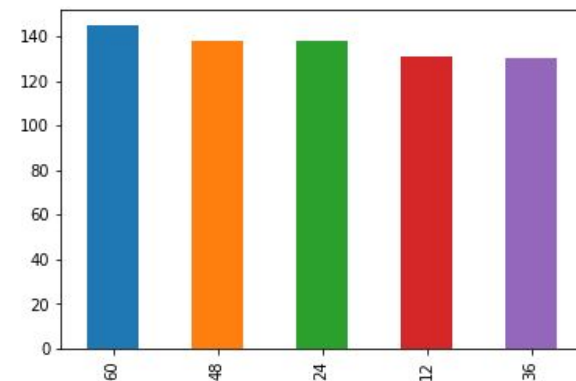
mean 151410.175953  
std 113372.406310  
min 4980.000000  
25% 66732.000000  
50% 116928.000000  
75% 210654.000000  
max 590820.000000

monthly payments



mean 4190.664223  
std 2215.830344  
min 304.000000  
25% 2477.000000  
50% 3934.000000  
75% 5813.500000  
max 9910.000000

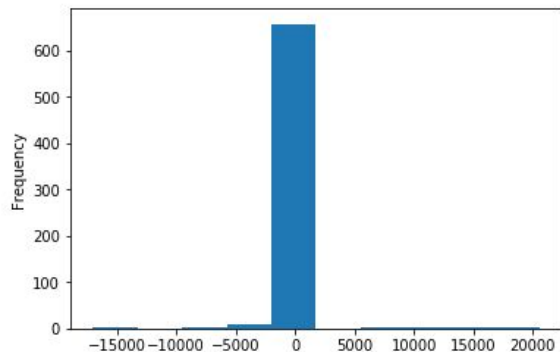
duration



60 : 145  
48 : 138  
24 : 138  
12 : 131  
36 : 130

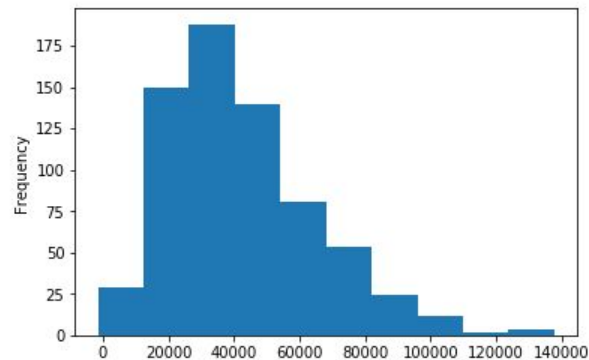
## Transactions

### Min balance



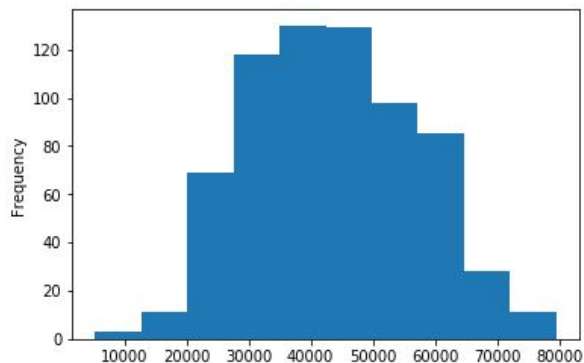
mean 651.994135  
std 1848.990442  
min -17030.000000  
25% 300.000000  
50% 600.000000  
75% 900.000000  
max 20574.000000

### Last balance



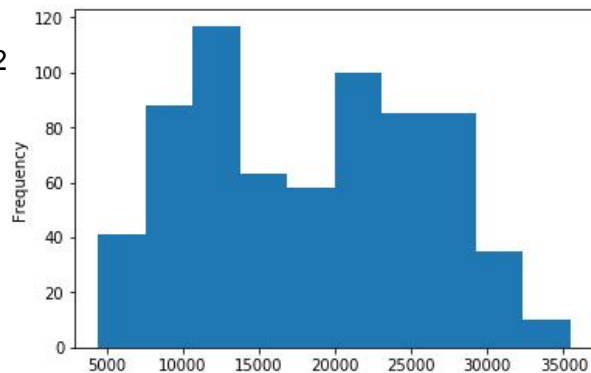
mean 42383.570381  
std 22516.595485  
min -1572.000000  
25% 25918.500000  
50% 38315.500000  
75% 54580.000000  
max 137843.000000

### Mean balance



mean 43610.168682  
std 13472.373437  
min 5350.000000  
25% 32833.081400  
50% 43207.928547  
75% 53614.015301  
max 79500.529762

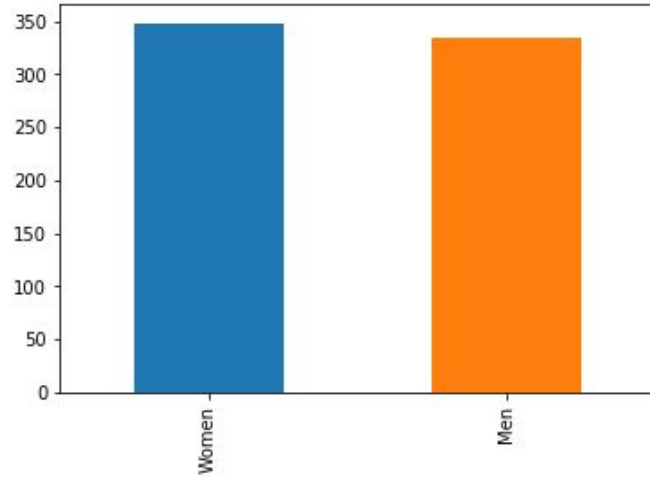
### std balance



mean 18355.398745  
std 7481.459342  
min 4462.706158  
25% 11687.797057  
50% 18893.263292  
75% 24705.349918  
max 35435.984338

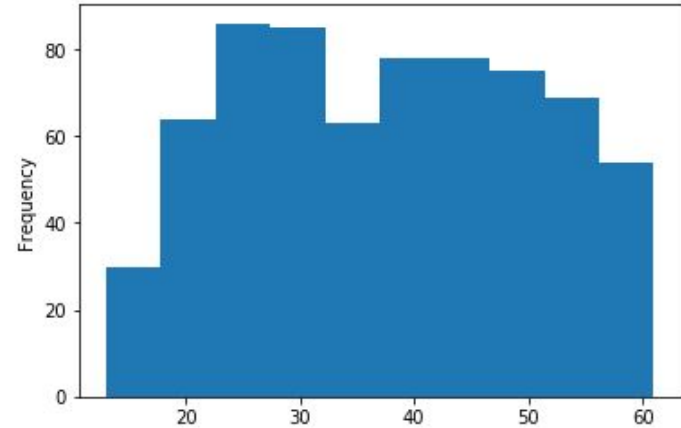
## Client

### Gender



Women 348  
Men 334

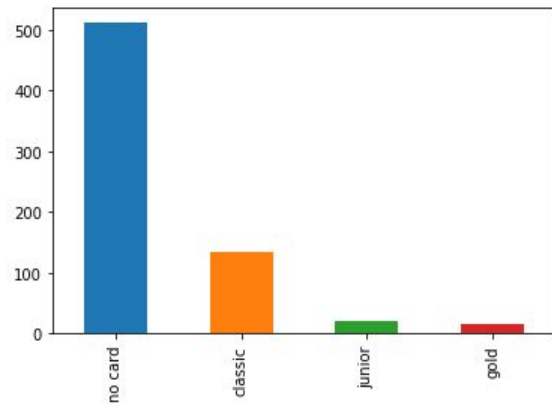
### Age



mean 37.630499  
std 12.768289  
min 13.000000  
25% 27.000000  
50% 37.000000  
75% 48.000000  
max 61.000000

## Credit Card

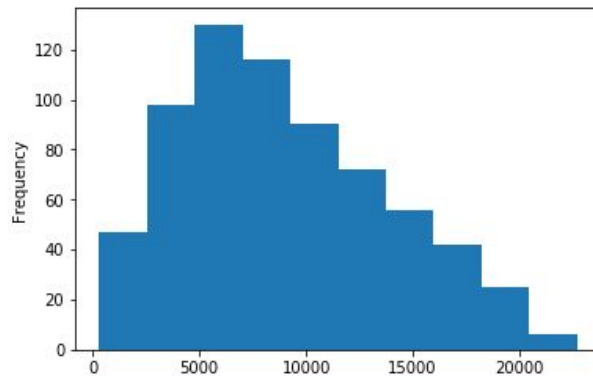
Card type



no_card	646
classic	28
junior	5
gold	3

## Orders

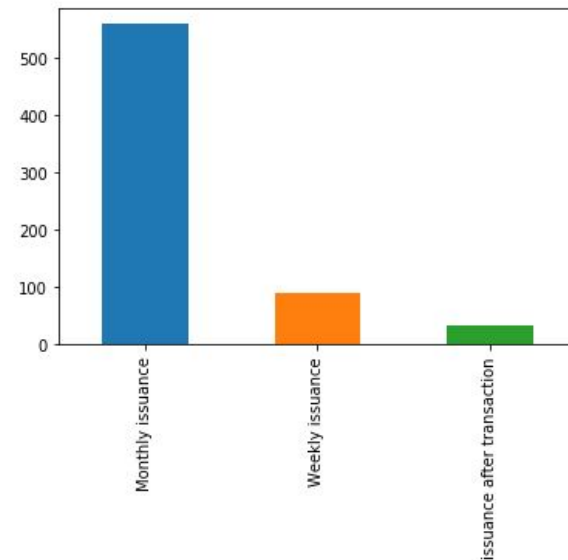
Debited\_amount



mean	9003.368035
std	4842.862438
min	312.000000
25%	5302.500000
50%	8047.500000
75%	12434.750000
max	22704.000000

## Account

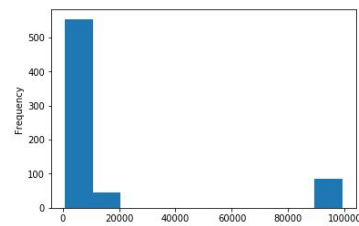
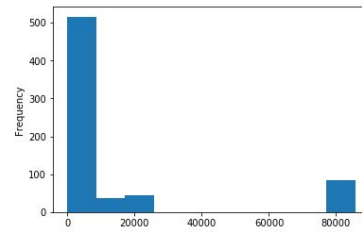
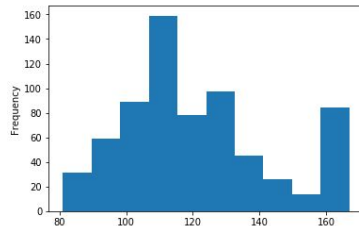
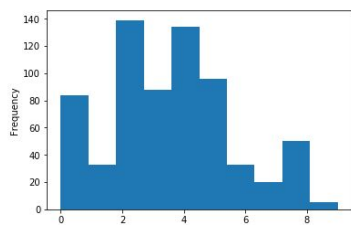
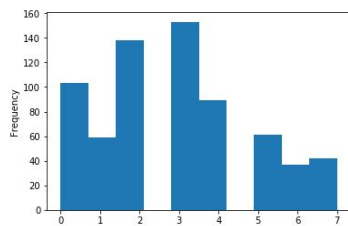
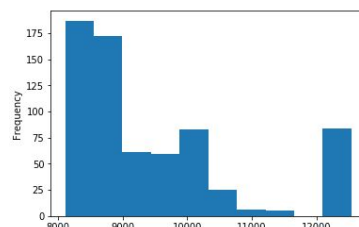
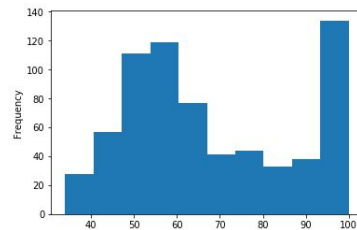
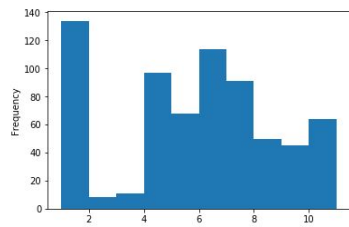
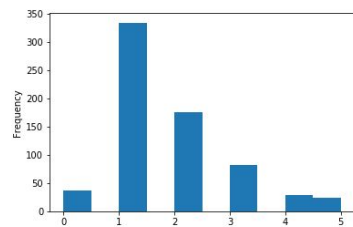
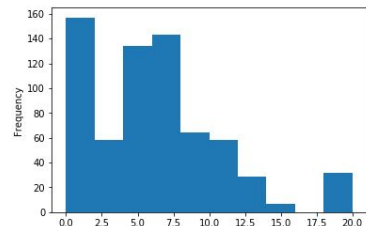
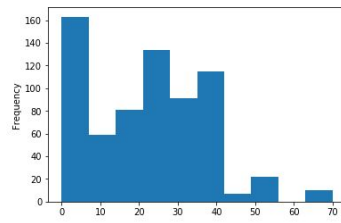
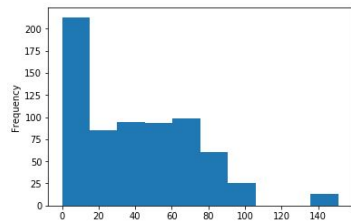
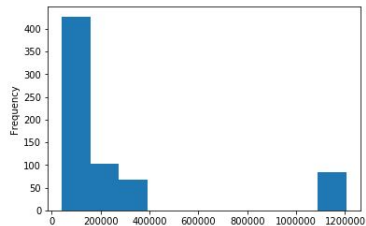
Frequency



Monthly issuance	559
Weekly issuance	91
issuance after transaction	32



## Demographic Data



A12

A13

A14

A15

A16

A4 : no. Of inhabitants

A5 : no. of municipalities with inhabitants < 499

A6 : no. of municipalities with inhabitants 500-1999

A7 : no. of municipalities with inhabitants 2000-9999

A8 : no. of municipalities with inhabitants >10000

A9 : no. of cities

A10 : ratio of urban inhabitants

A11 : average salary

A12 : unemployment rate '95

A13 : unemployment rate '96

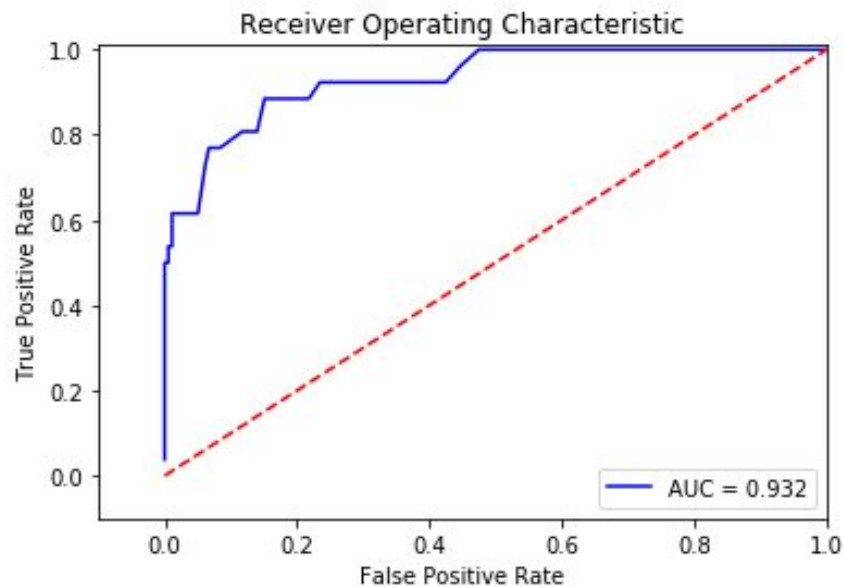
A14 : no. of entrepreneurs per 1000 inhabitants

A15 : no. of committed crimes '95

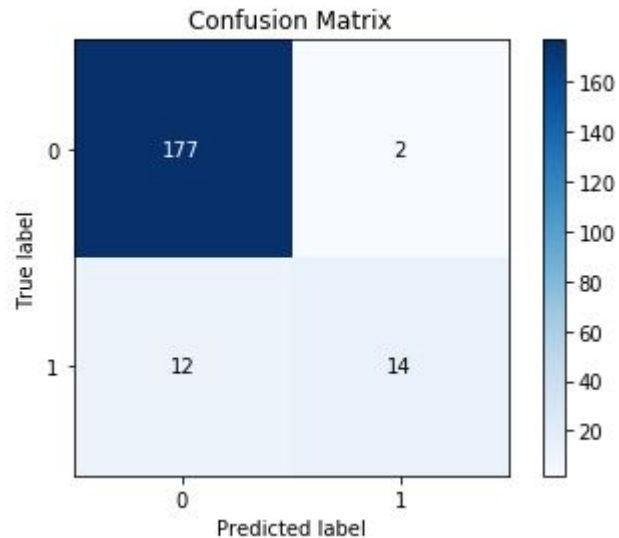
A16 : no. of committed crimes '96

# ML Model

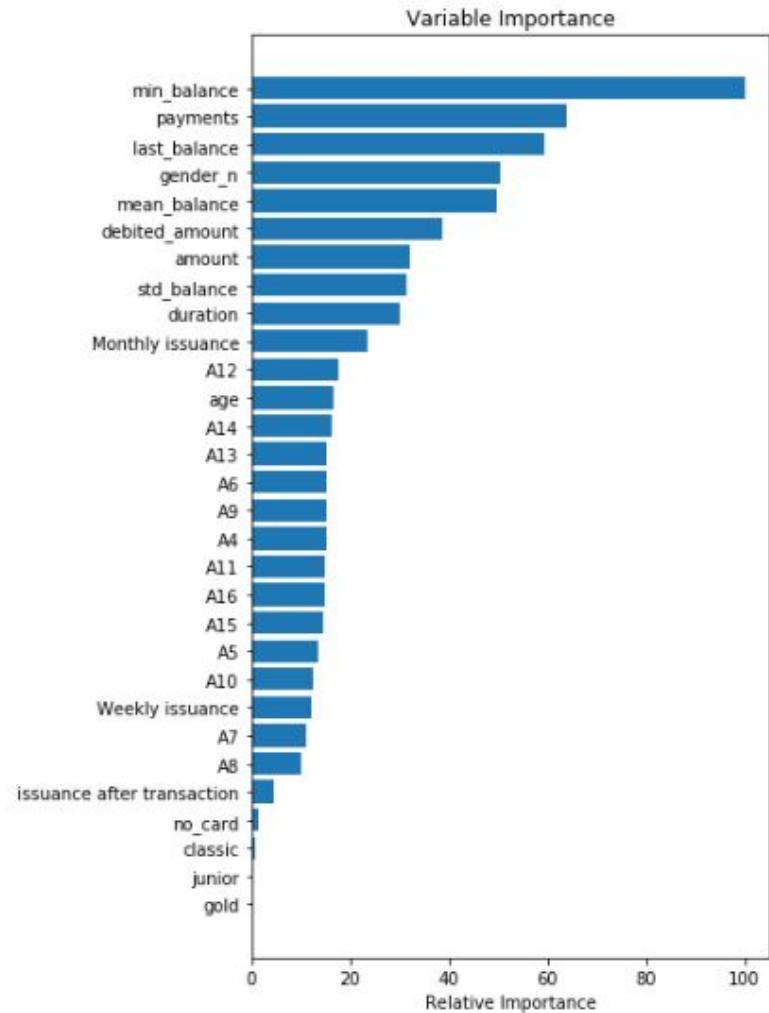
## 1. RandomForest



	precision	recall	f1-score	support
0	0.94	0.99	0.96	179
1	0.88	0.54	0.67	26
avg / total	0.93	0.93	0.92	205

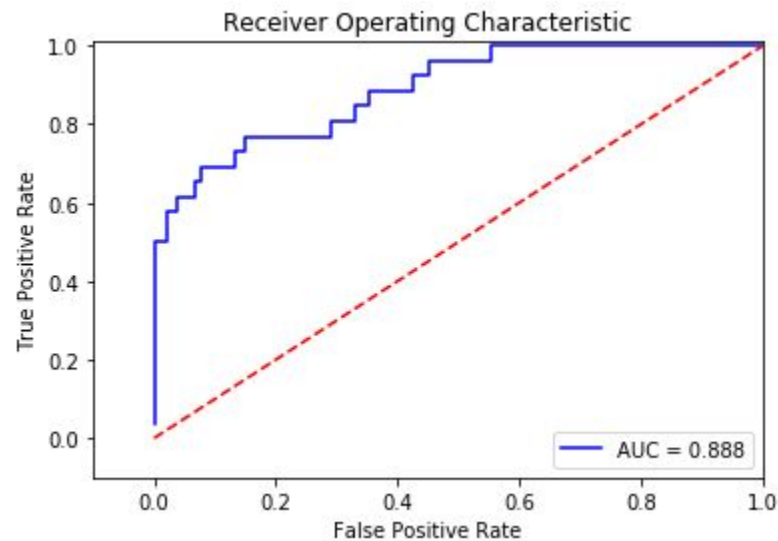


# RandomForest Variable Importance

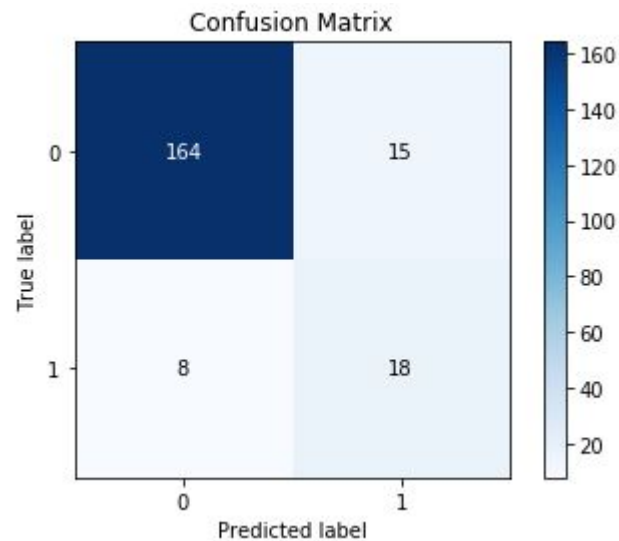


# ML Model

## 2. AdaBoost

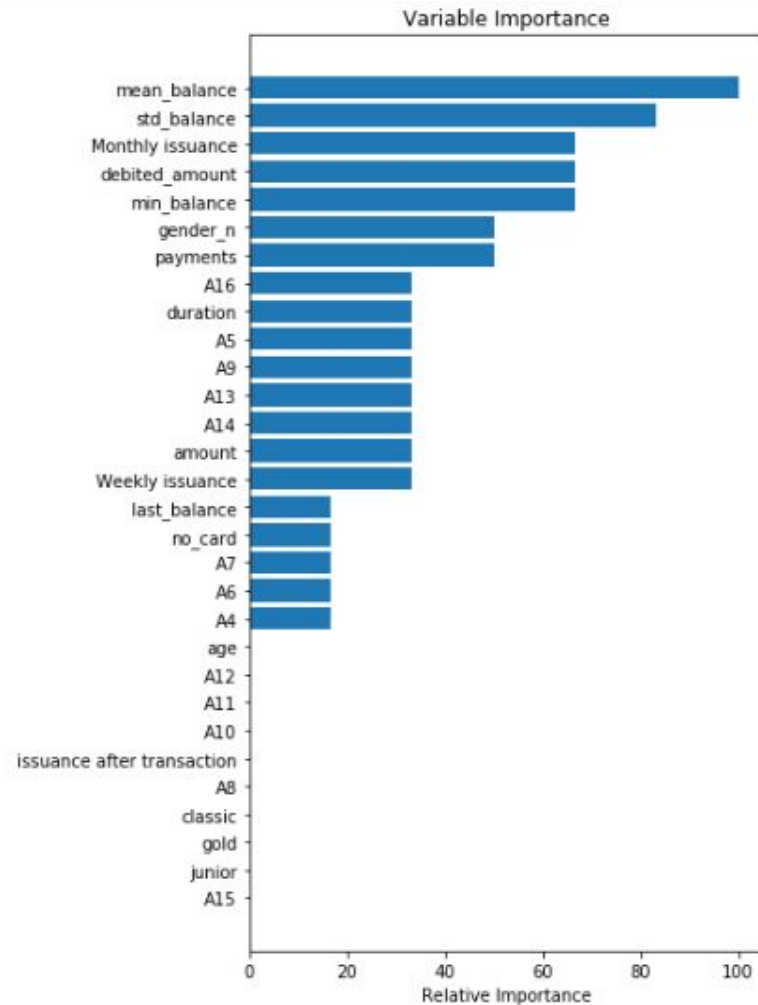


	precision	recall	f1-score	support
0	0.95	0.92	0.93	179
1	0.55	0.69	0.61	26
avg / total	0.90	0.89	0.89	205



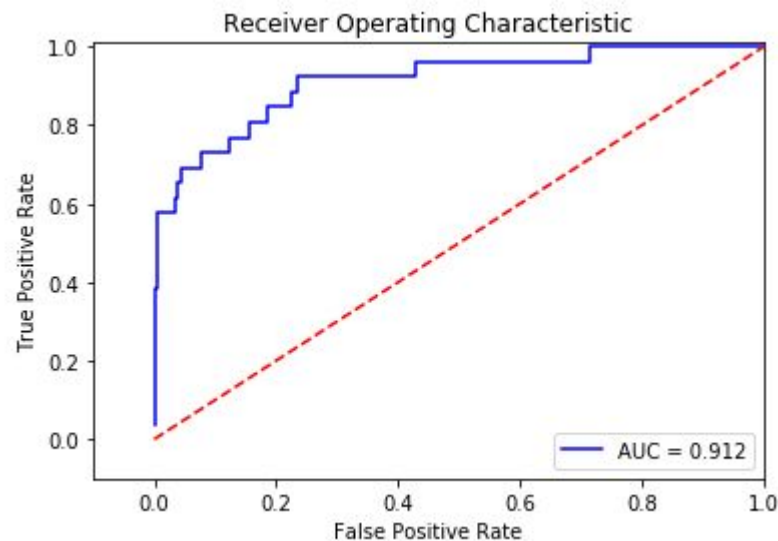
# Adaboost

## Variable Importance

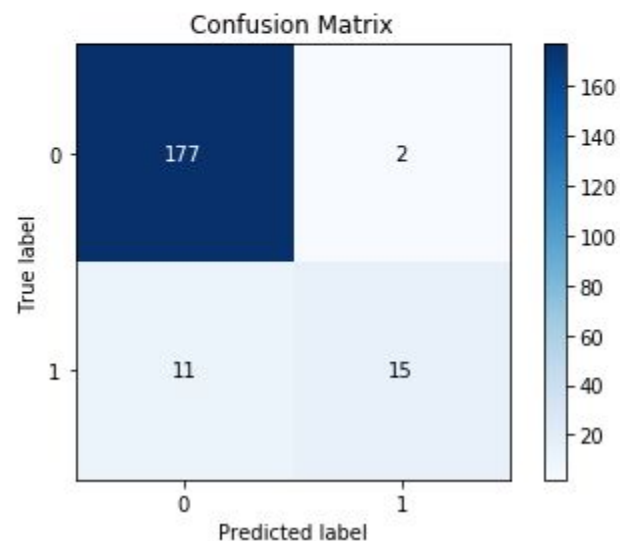


# ML Model

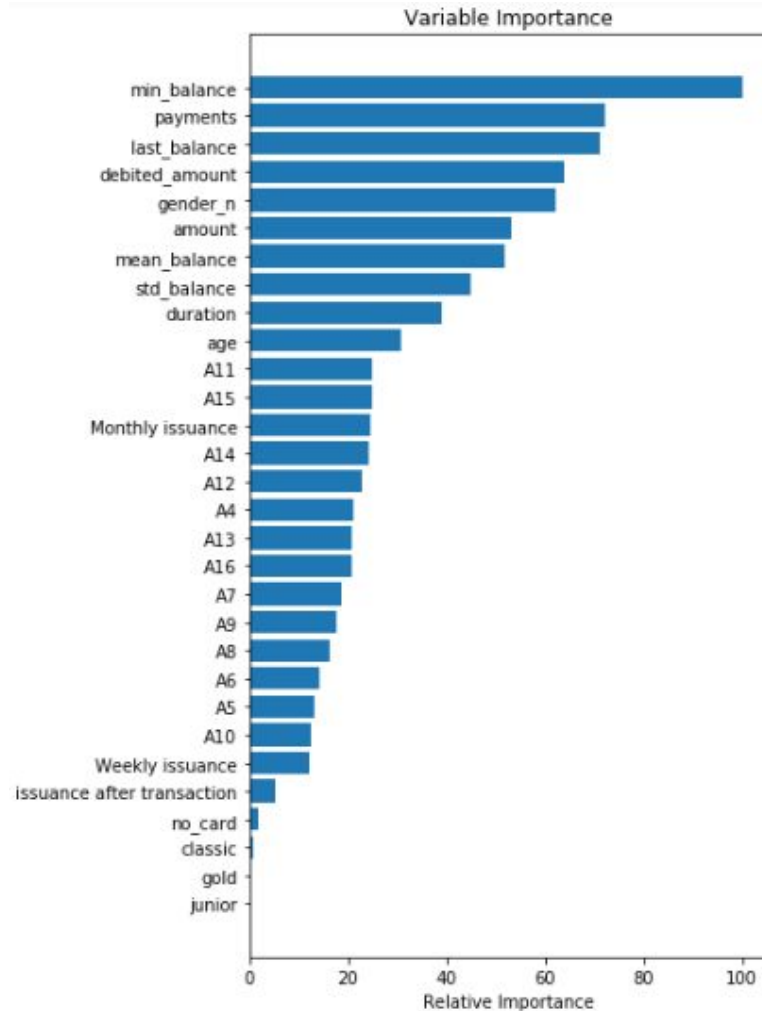
## 3. Gradient Boost



	precision	recall	f1-score	support
0	0.94	0.99	0.96	179
1	0.88	0.58	0.70	26
avg / total	0.93	0.94	0.93	205



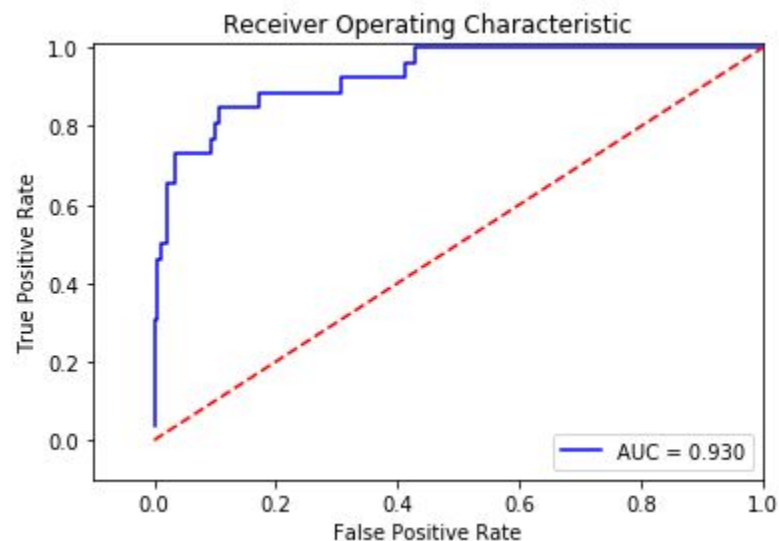
# Gradient Boost Variable Importance



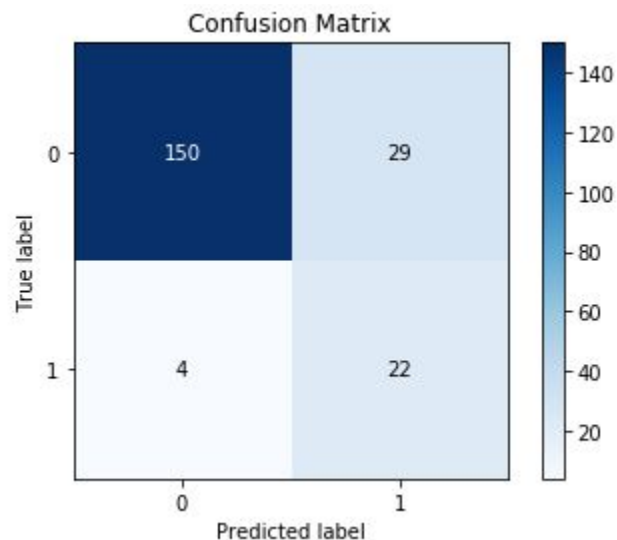


# ML Model

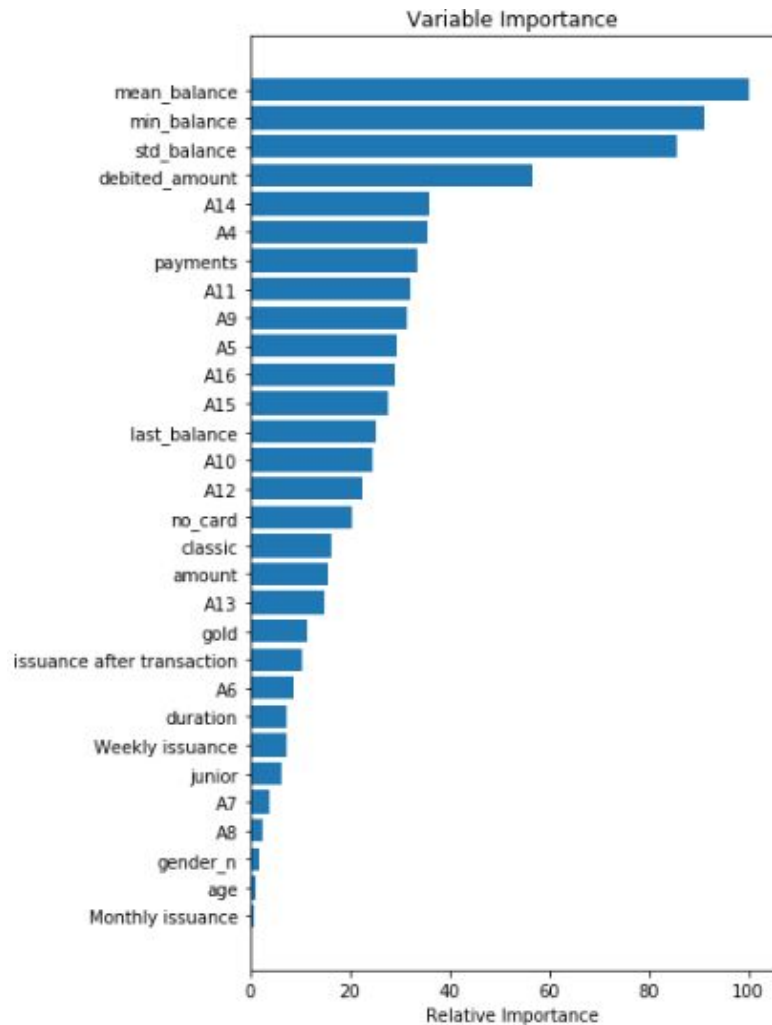
## 4. Logistic Regression



	precision	recall	f1-score	support
0	0.97	0.84	0.90	179
1	0.43	0.85	0.57	26
avg / total	0.91	0.84	0.86	205



# Logistic Regression Variable Importance



# ML Model Summary

	Logistic Regression	Random Forest	AdaBoost	Gradient Boost
AUC	0.93	0.932	0.888	0.912
Recall for default	0.85	0.54	0.69	0.58
Top 5 Key Predictors	Mean balance Min balance Std balance Debited amount A14	Min balance Payments Last balance Gender Mean balance	Mean balance Std balance Monthly Issuance Debited amount Min balance	Min balance Payment Last balance Debited amount Gender

A14 : no. of entrepreneurs per 1000 inhabitants