

Predicting Credit Card Default Using Machine Learning- Technical Presentation

Presenter: Didem Bulut Aykurt | Mentor: Akshay Jhawar | Springboard Data Analyst Career Track
Date: 01/29/2021

This dataset contains information on default payments, demographic factors, credit data, history of payment, and bill statements of credit card clients in Taiwan from April 2005 to September 2005.

1) lower credit limits are more likely to default

Younger customers have lower limits credit, and they don't default as much as the older customers.

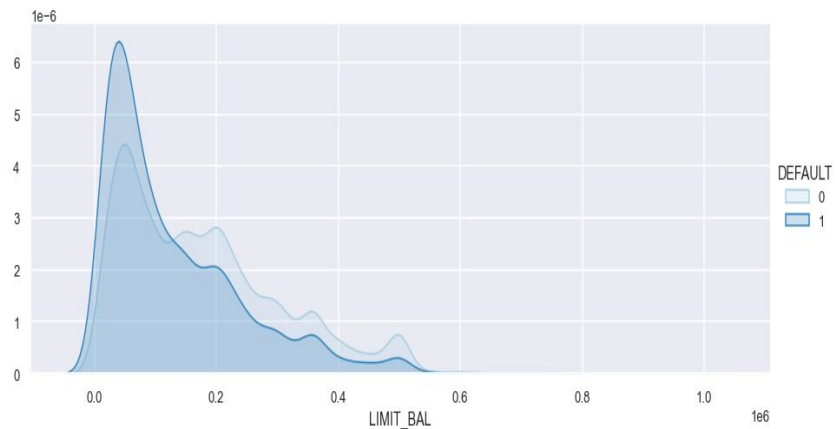
2) Each month, paid minimum only(0) has a lower chance default, but late 2 month(2) repayment tend to default.

When payment is delayed more than 2 months, the chances of default goes higher than 50%.

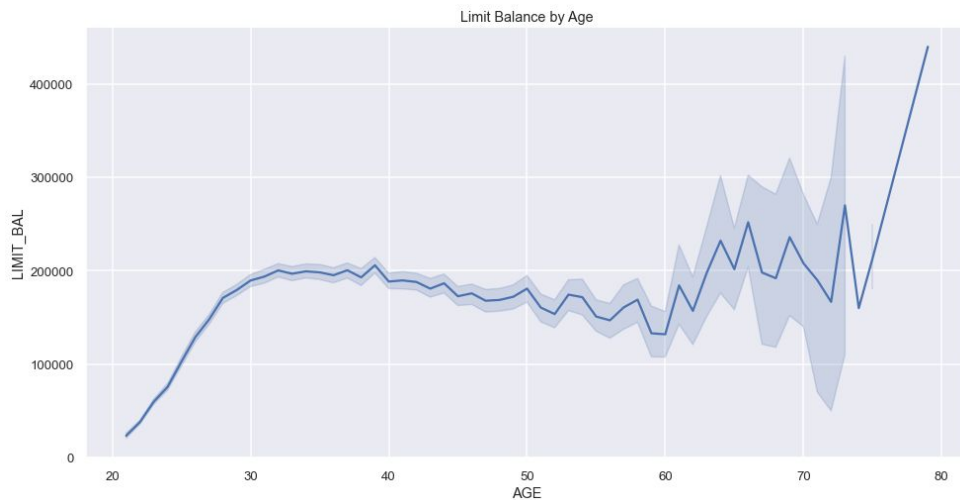
3)When credit limit is 500T, default sharply increase and billing amount of median increase.

Highest Bill Amount above 1M and lowest one approximately -2T from September 2005.

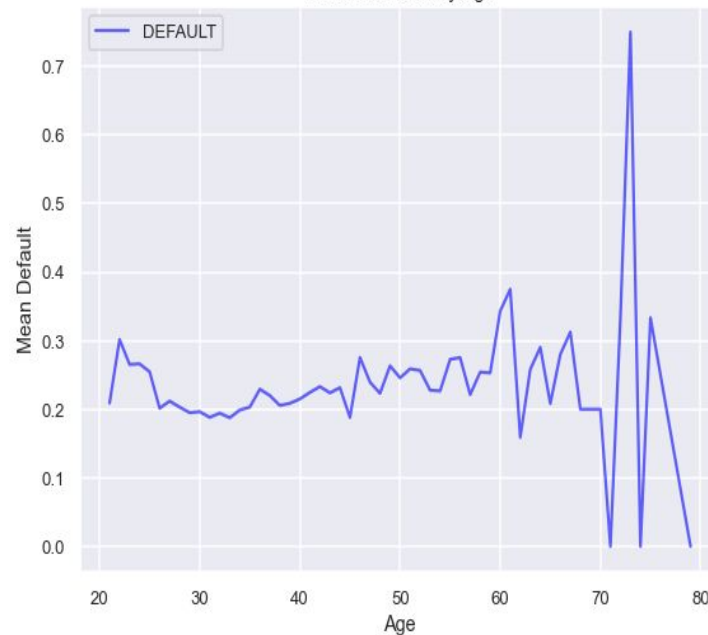
Limit Balance and Default Rate Distribution



From the charts below we see that the default rate was markedly lower for people under 60 compared to people above 60 years old. The default rate was around 30 % for people above 60 years old and only about 20 % for people below 60. As credit limit also higher for above 60 years old.



Percent Default by Age



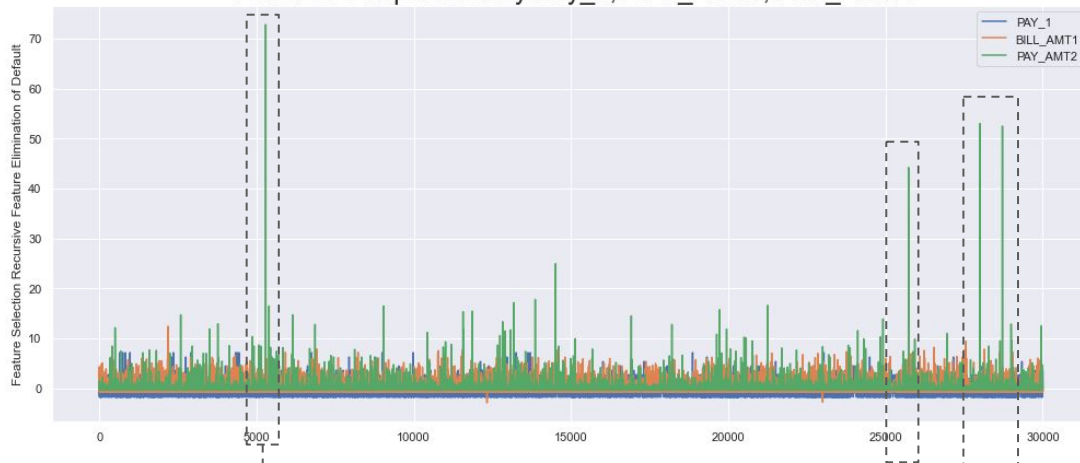
Through a combination of visual and statistical methodologies, there (3) key variables were associated with DEFAULT parment that enabled a high-level equation that effectively modelled and pick up Default payment behaviour.

Visual Methodologies



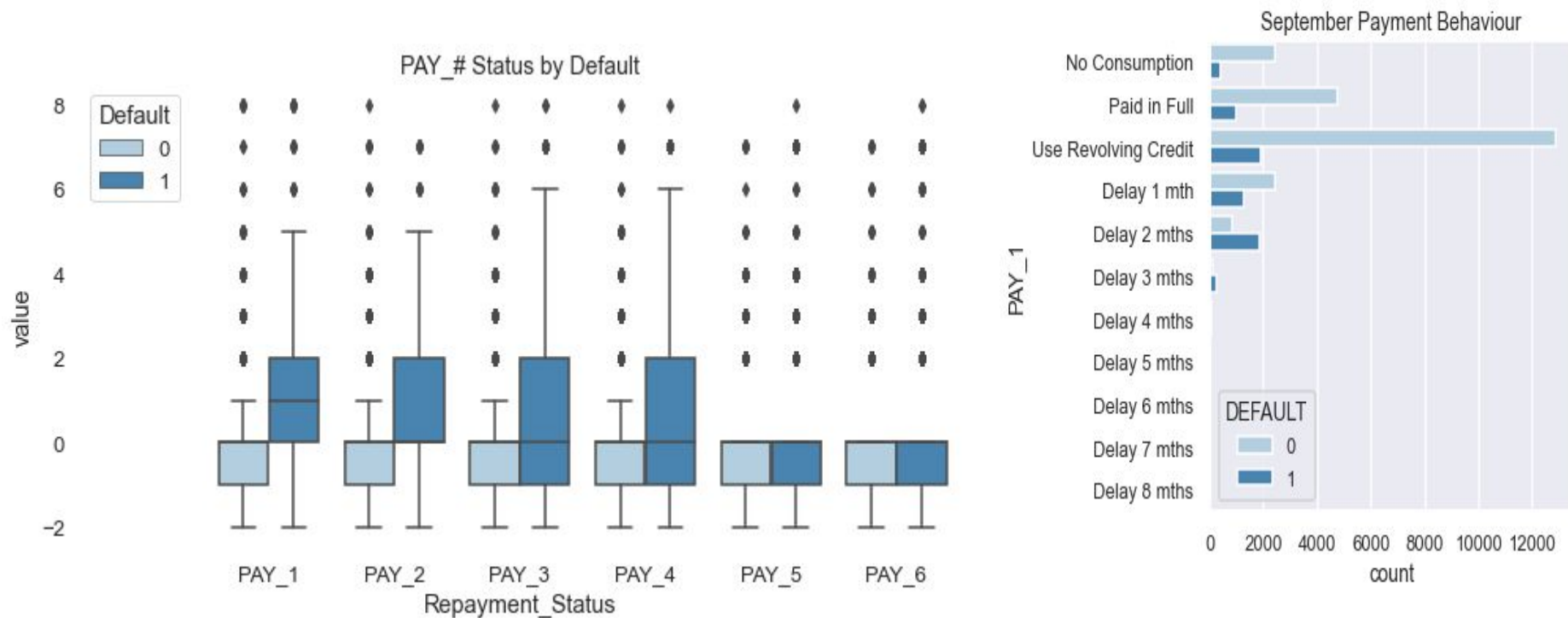
Statistical Methodologies

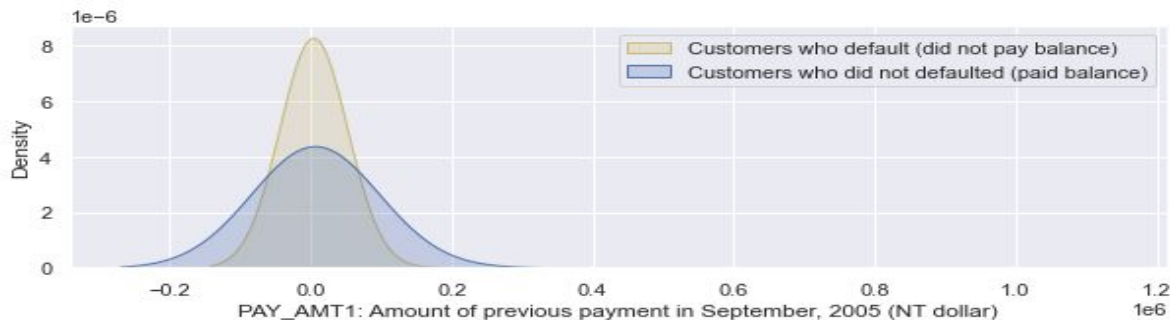
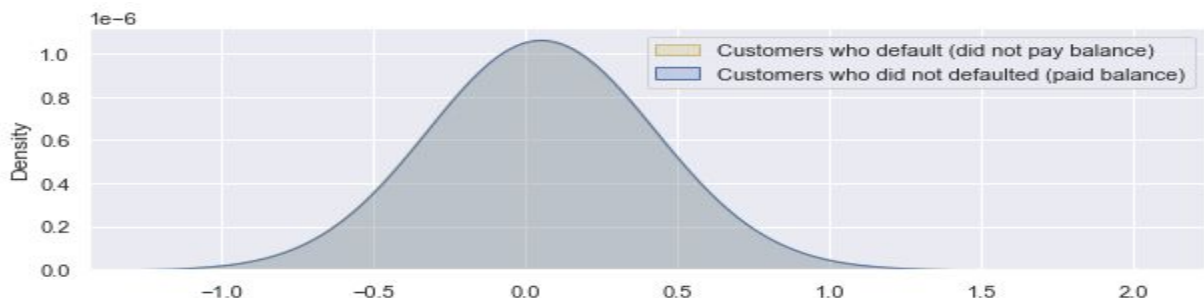
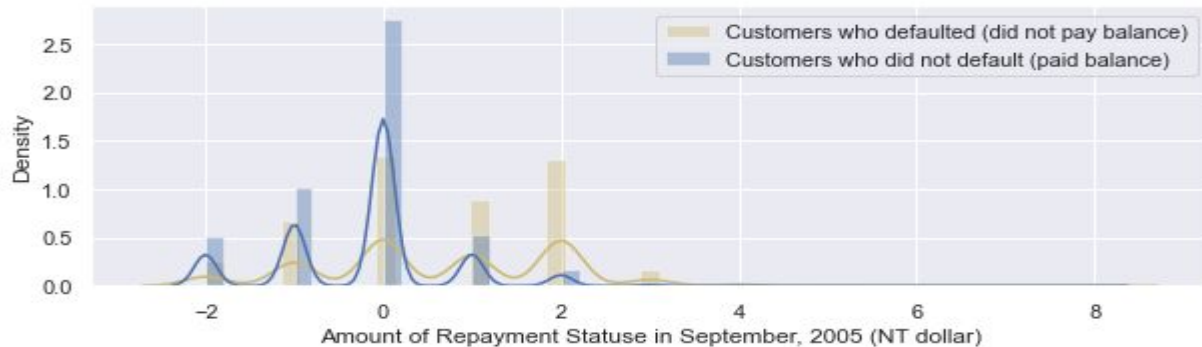
Features 3 Important Key Pay_1, BILL_AMT1, PAY_AMT2



4 highest peak for Amount of previous payment in August.

It seems that PAY_1 (Repayment status in September) and PAY_2 (Repayment status in August) have more discriminatory power the repayment status in other months. Those Using Revolving Credit (paid only minimum) and those delayed for 2 months have the highest Default Count.



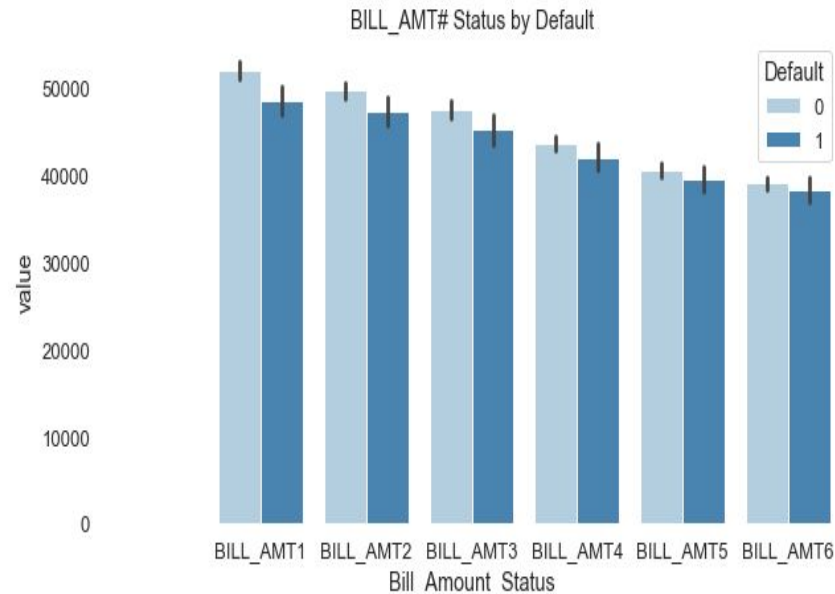
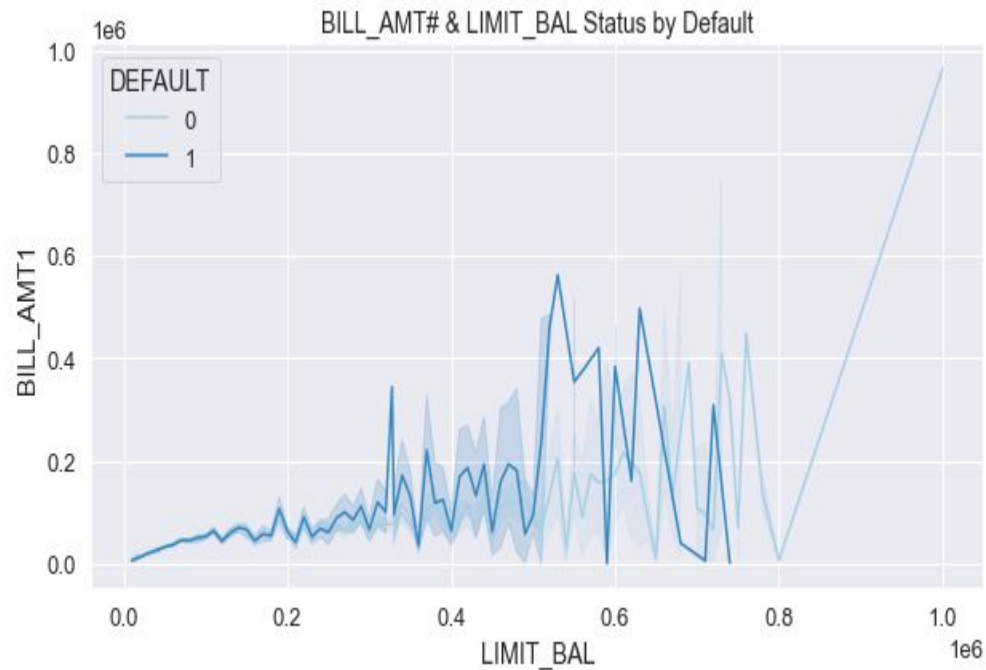


First plot about 'PAY_1' repayment status; there are 6 peak and highest peak destiny on-time repayment status ('PAY_1'= 0 with 'DEFAULT'=0). First 3 peak has a highest repayment after the 0=on-time peak exchanged highest is default lowest one is nondefault.

-Second, amount of bill statement 'BILL_AMT#'; there is a one peak, and default and nondefault variables are closely same.

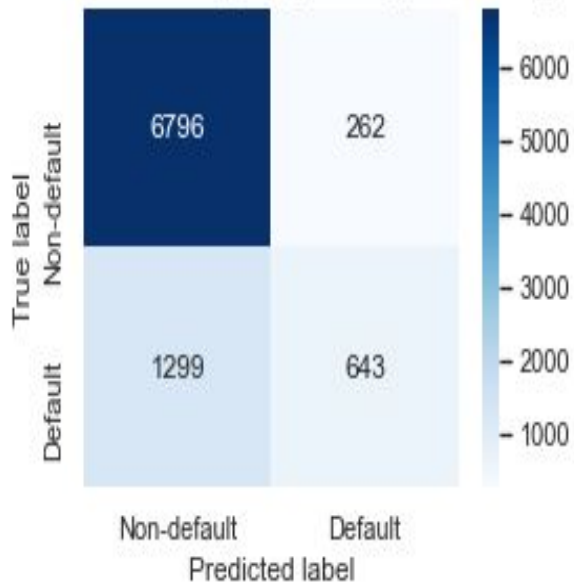
-Third, amount of previous payment 'PAY_AMT#'; there is a one peak, and default destiny higher than nondefault payment.

'**BILL_AMT**' variables have a strong positive correlation themselves and LIMIT_BAL, but there is a weak correlation between 'DEFAULT'. When credit limit is 500 Default and Bill amount of median sharply increase.



Repayment Status 'PAY_#', which shows their delay status, is the most important indicator for Default. Important for feature and accuracy 83%, recall 0.33.

Confusion Matrix - Single Logistic Regression PAY_1



Variables correlated with DEFAULT

