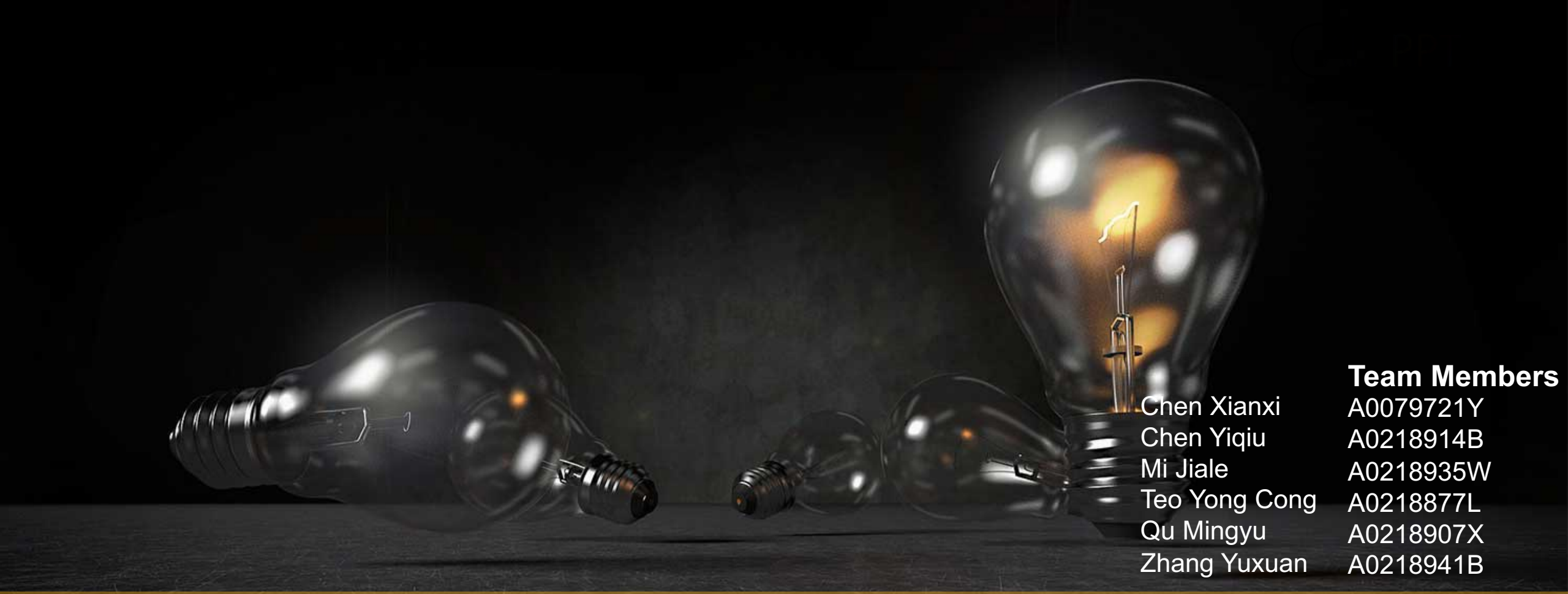


Team Members	
Chen Xianxi	A0079721Y
Chen Yiqiu	A0218914B
Mi Jiale	A0218935W
Teo Yong Cong	A0218877L
Qu Mingyu	A0218907X
Zhang Yuxuan	A0218941B

DBA 5102 – Business Analytics Capstone Module

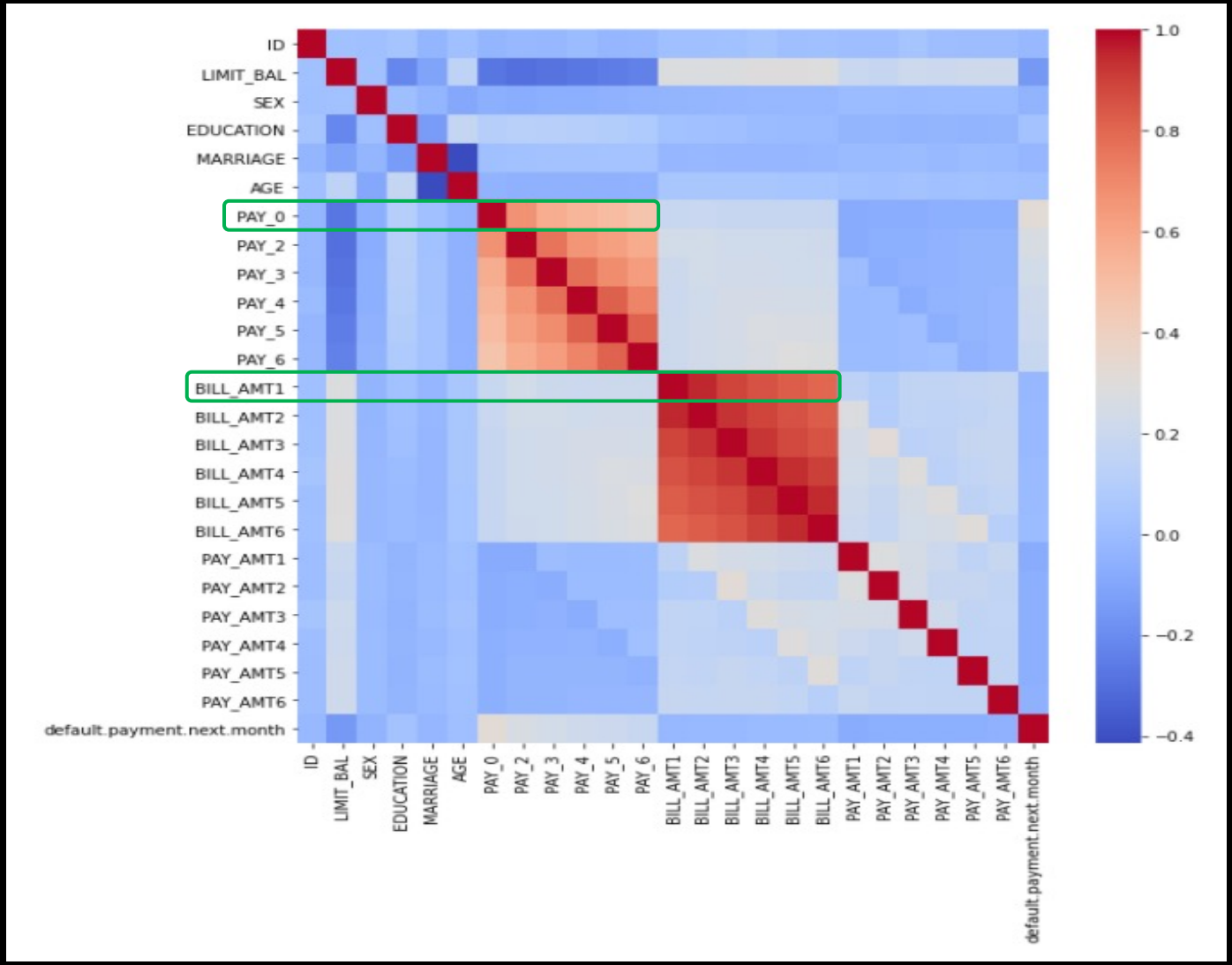
Machine Learning Interpretability Model for Default Prediction

Group Explanandum

Aim

Build a *highly explainable* credit card payment default prediction model

Exploratory Data Analysis

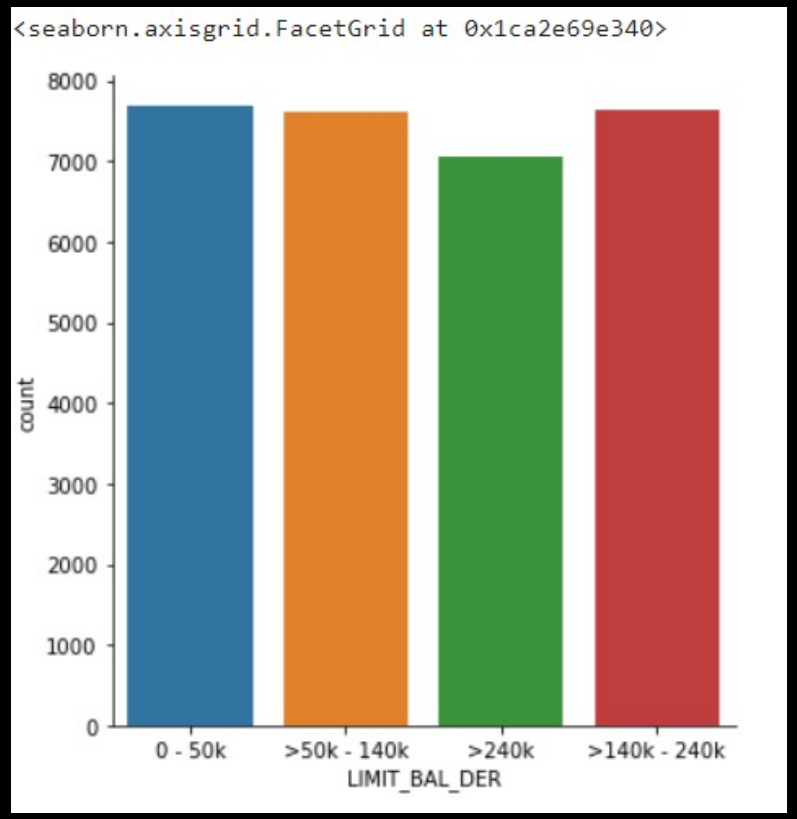


- From the correlation heatmap, the clients' demographics are generally uncorrelated while, expectedly, repayment status and bill amount are fairly correlated across the 6 months. For example:
  - PAY_0 is correlated with PAY_2 to PAY_6
  - BILL_AMT1 is correlated with BILL_AMT2 to BILL_AMT6

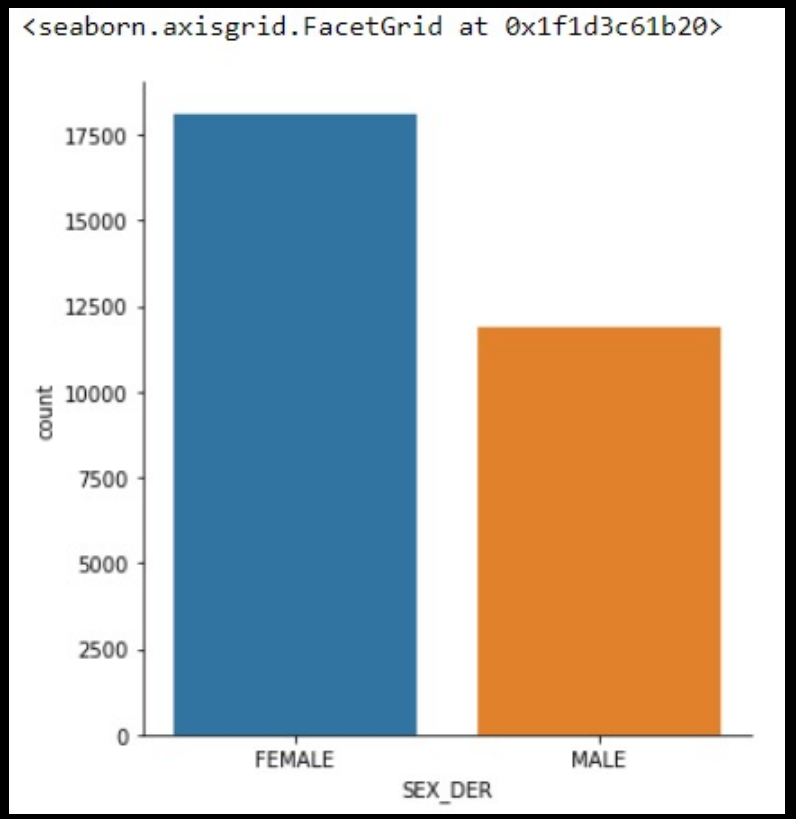
Exploratory Data Analysis

Dataset

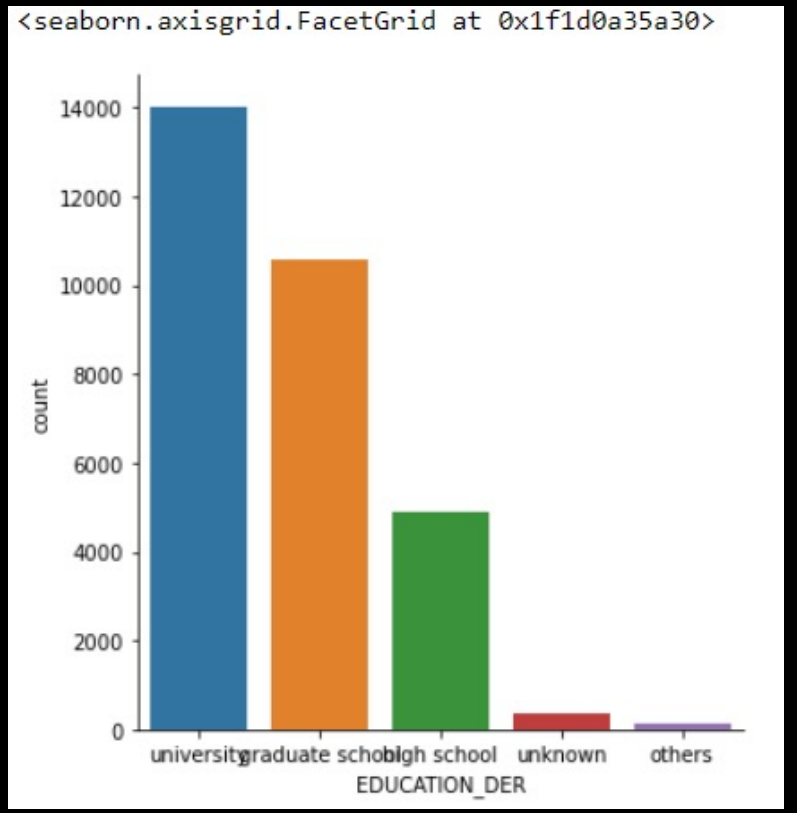
Distribution of demographic factors, credit limit data, history of payment of credit card clients of a Taiwanese bank from Apr to Sep 2005



Derived Limit Balance



Derived Sex

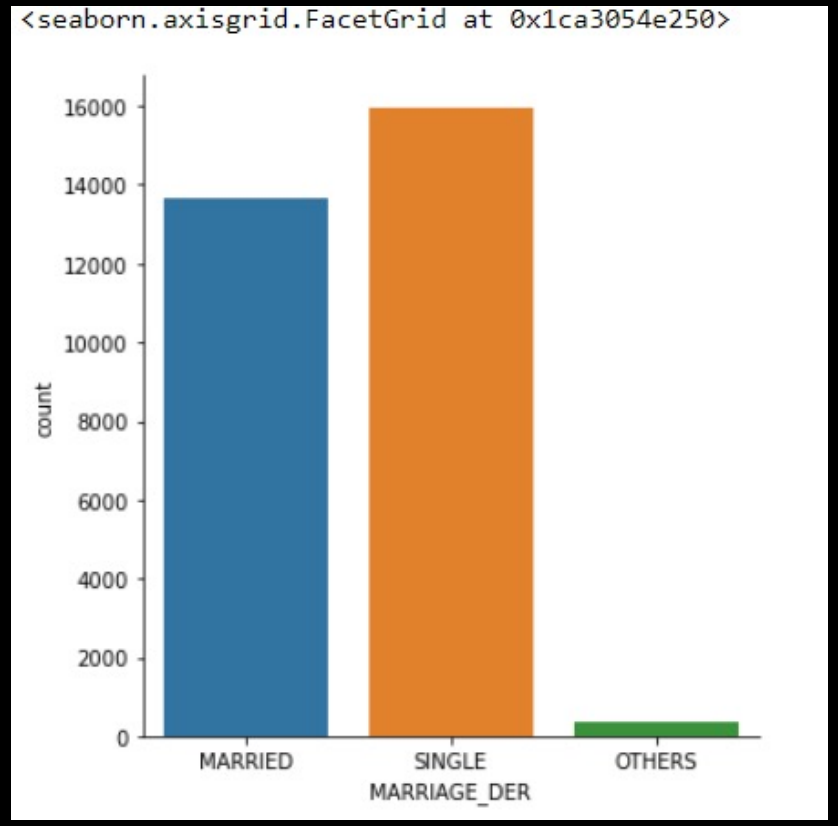


Derived Education

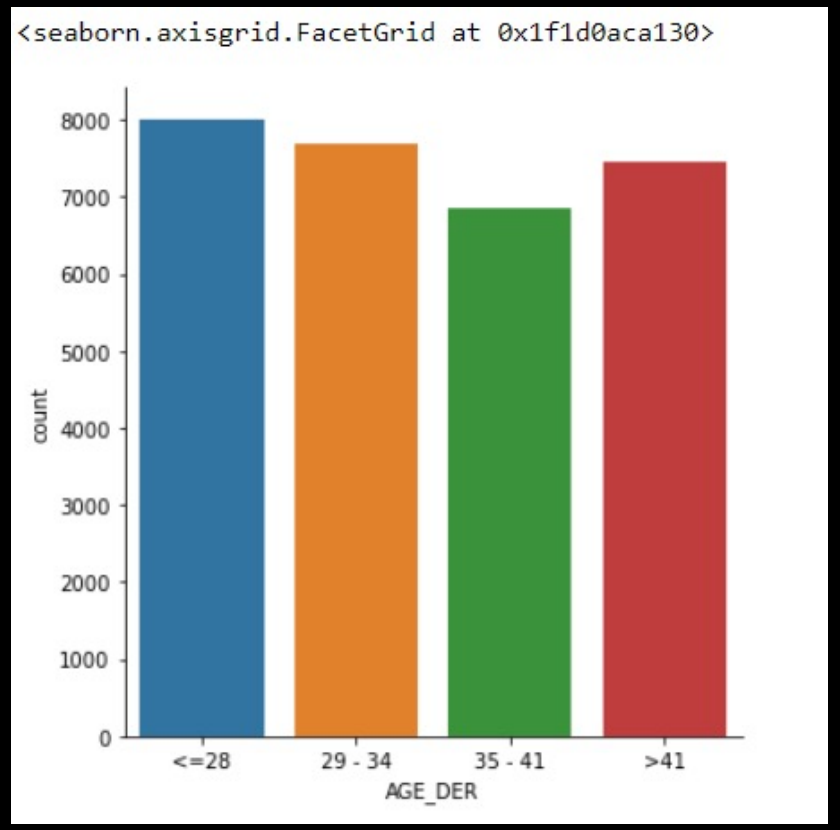
Exploratory Data Analysis

Dataset

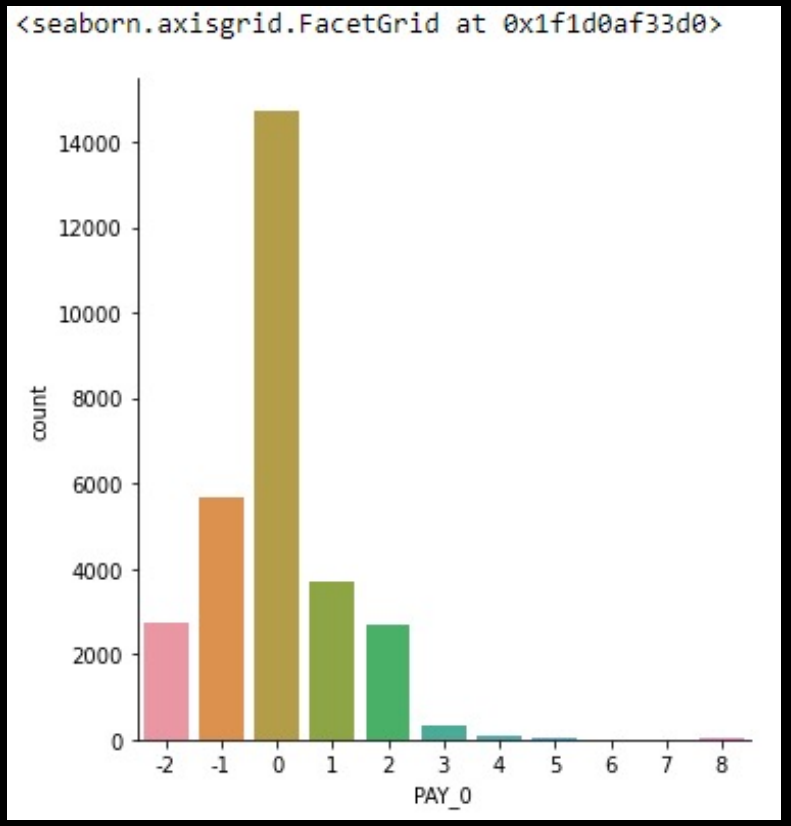
Distribution of demographic factors, credit limit data, history of payment of credit card clients of a Taiwanese bank from Apr to Sep 2005



Derived Marriage



Derived Age

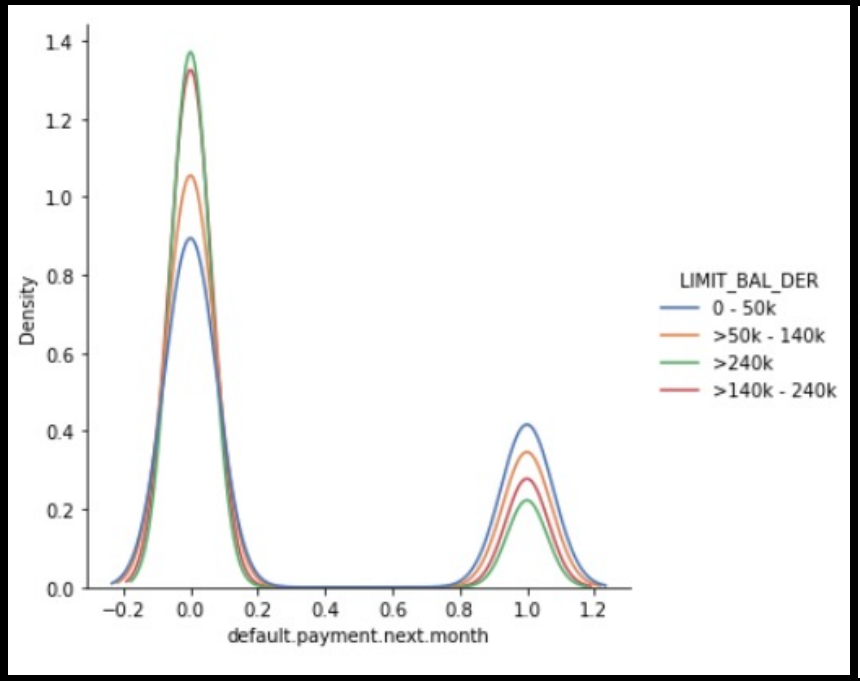


Pay_0

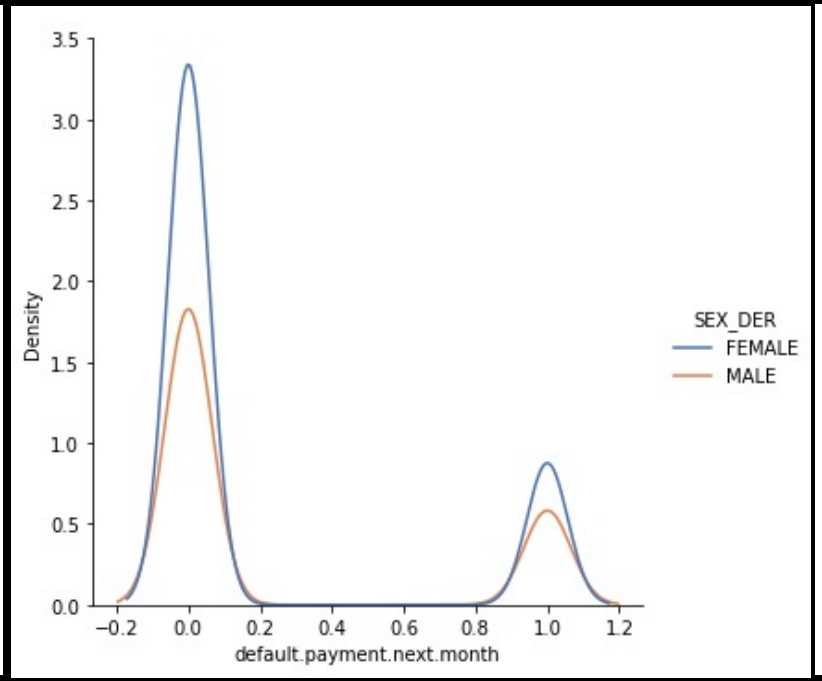
Exploratory Data Analysis

Dataset

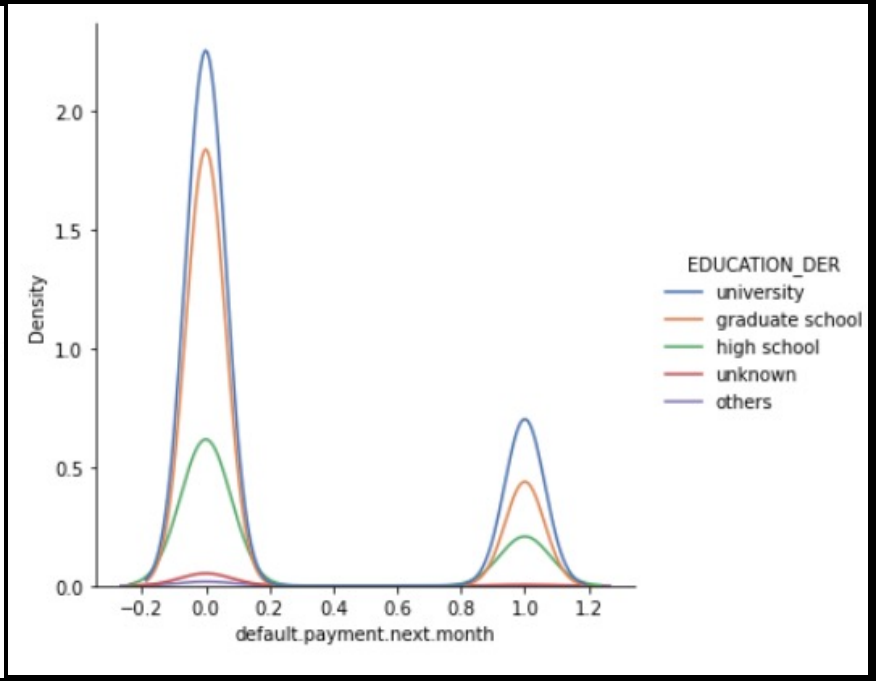
Information on default against data factors of credit card clients of a Taiwanese bank from Apr to Sep 2005



Default against Limit Balance



Default against Sex

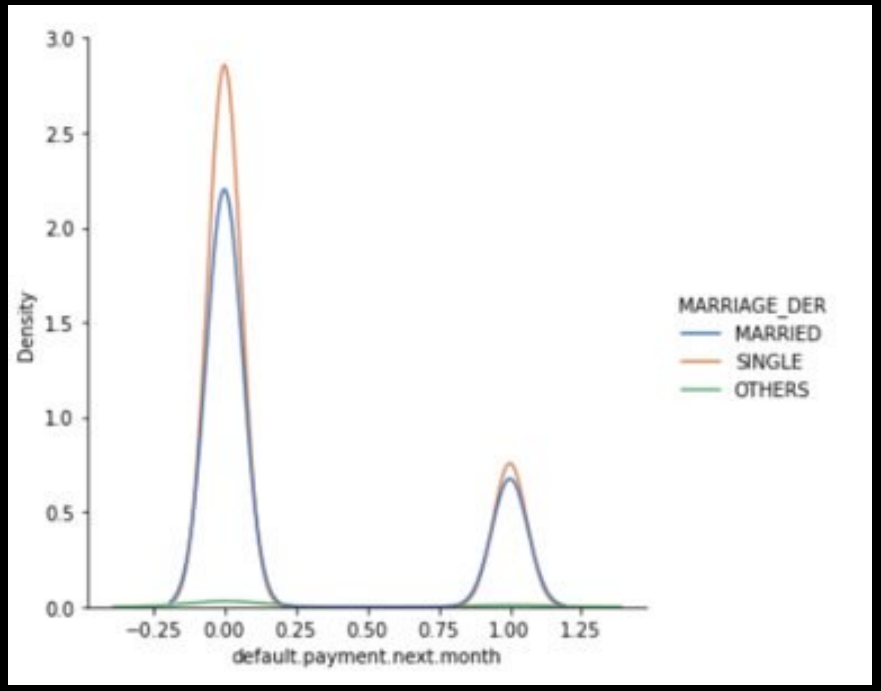


Default against Education

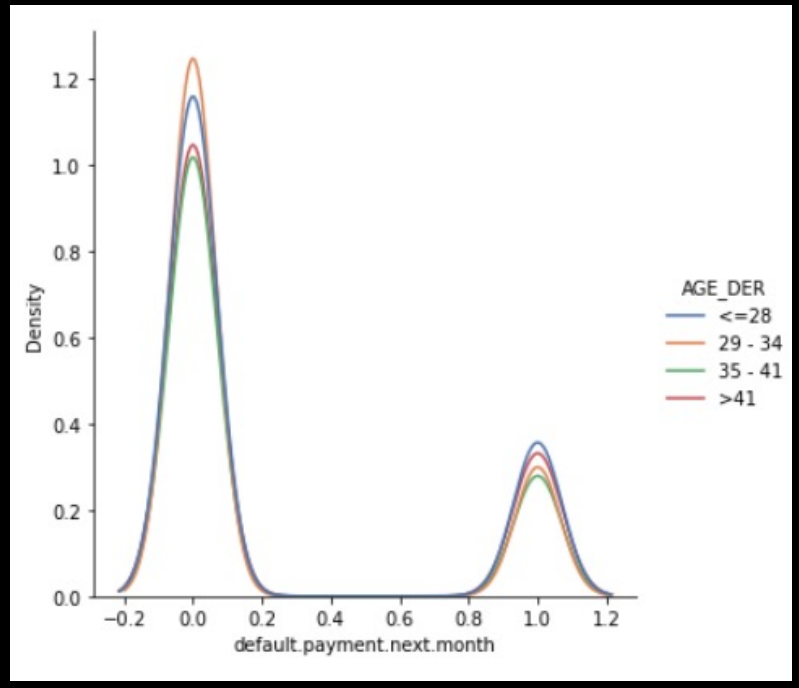
Exploratory Data Analysis

Dataset

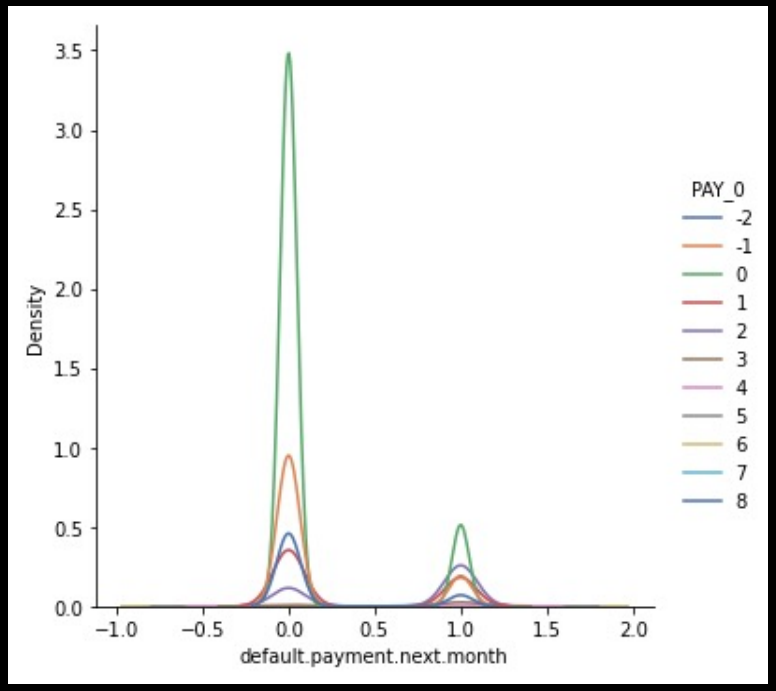
Information on default against data factors of credit card clients of a Taiwanese bank from Apr to Sep 2005



Default against Marraige



Default against Age



Default against Pay_0

Data Pre-Processing

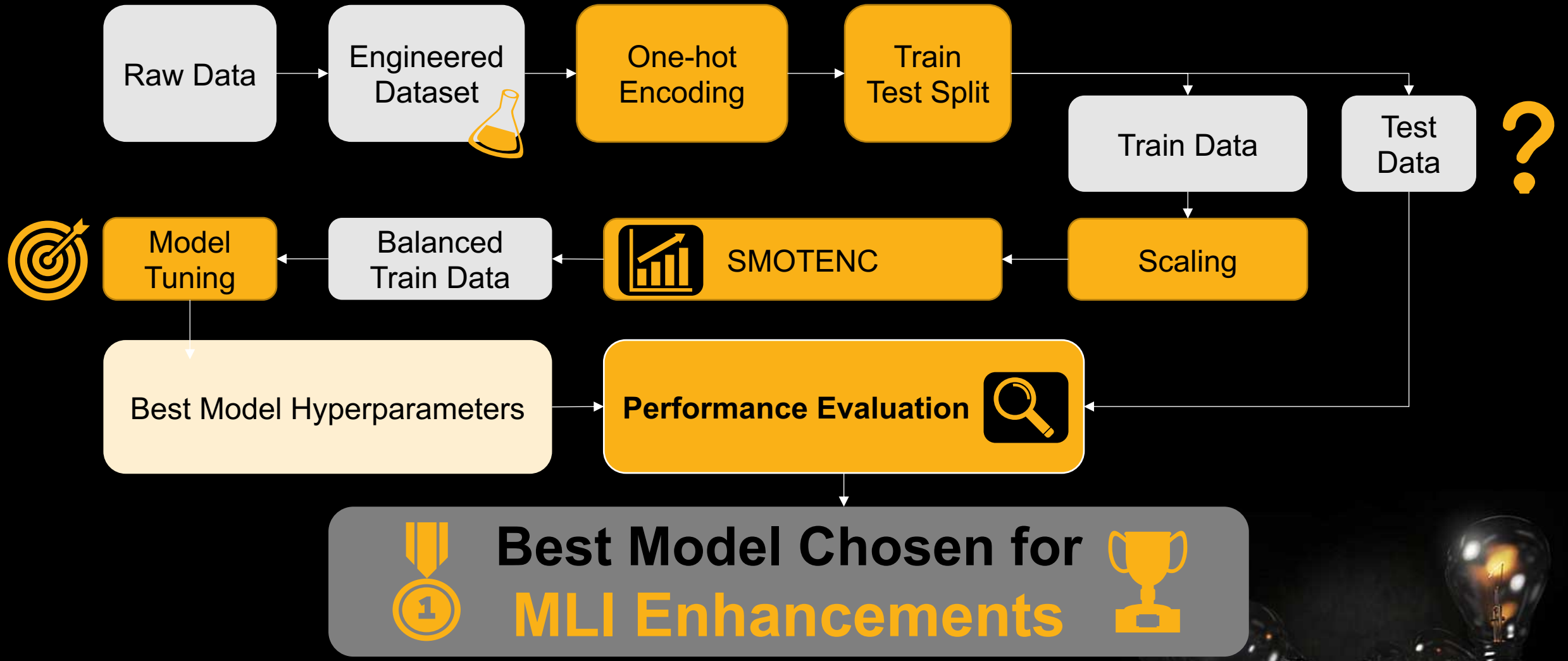
S/No.	Variable	Description	Type	One-hot Encoding?
1	ID	Client ID	Categorical	No – dropped
2	LIMIT_BAL	Amount of given credit in NT dollars	Continuous	No
3	SEX	Client gender	Categorical	Yes
4	EDUCATION	Client education level	Categorical	Yes (merged classes 4, 5 and 6)
5	MARRIAGE	Marital status	Categorical	Yes
6	AGE	Client age in years	Categorical	Yes (encoded age bands)
7	PAY_0 to PAY_6	Repayment status for months Apr to Sep 2005	Categorical	Yes
8	BILL_AMT1 to BILL_AMT6	Amount of bill statement in NT dollar for months Apr to Sep 2005	Continuous	No
9	PAY_AMT1 – PAY_AMT6	Amount of payment in NT dollar for previous month for months Apr to Sep 2005	Continuous	No
10	default.payment.next. month	Whether the client defaulted on payment in prediction period (i.e. Oct 2005)	Categorical	No

Feature Engineering

- Created new features as shown in the table below

S/No.	Variable	Description	Base features	Type	One-hot Encoding?
1	Sex_Marriage	Combination of gender and marital status	SEX & MARRIAGE	Categorical	Yes
2	Sex_Education	Combination of gender and education level	SEX & EDUCATION	Categorical	Yes
3	Sex_Education_Marriage	Combination of gender, education level and marital status	SEX & EDUCATION & MARRIAGE	Categorical	Yes
4	Total_Bill_Amount	Summing up BILL_AMT1 to BILL_AMT6	BILL_AMT1 to BILL_AMT6	Continuous	No
5	Total_Payment_Amount	Summing up PAY_AMT1 – PAY_AMT6	PAY_AMT1 – PAY_AMT6	Continuous	No

Evaluation Methodology



Evaluation Metric

Evaluation Metric: Recall

$$recall = \frac{tp}{tp + fn}$$

- Selected ‘Recall’ as our evaluation metric as we want a metric that is more sensitive towards (i.e. minimise) false negative predictions
- This is valuable to banks because they need to set aside Specific Provisions (SP) in accordance with Basel Framework
- Using Recall as our evaluation metric would enable us to increase our accuracy in predicting probability of default in the portfolio and to adjust the SP budget accordingly

Models and Hyperparameter Tuning

Evaluation Metric: Recall

Algorithm	Hyperparameters
XG Boost	• n_estimators: 75, 100, 125, 150 • learning_rate: 1, 0.3, 0.1 • max_depth: 5, 6, 7
Random Forest	• n_estimators: 25, 50, 100 • max_features: 3, 5, 7, ‘auto’ • min_samples_split: 2, 5, 7 • criterion: ‘gini, ‘entropy’
Light GBM	• n_estimators: 25, 50, 75, 100 • num_leaves: 5, 10, 20, 25 • learning_rate: 0.1, 0.05, 0.01
Logistic Regression	• penalty: 'l2', 'elasticnet', 'l1' • C: 1.0, 0.5, 0.1, 0.05

Modeling

XG Boost	Random Forest	Light GBM ✓	Logistic Regression
0.64	0.67	**0.68**	0.57

Test Recall Scores

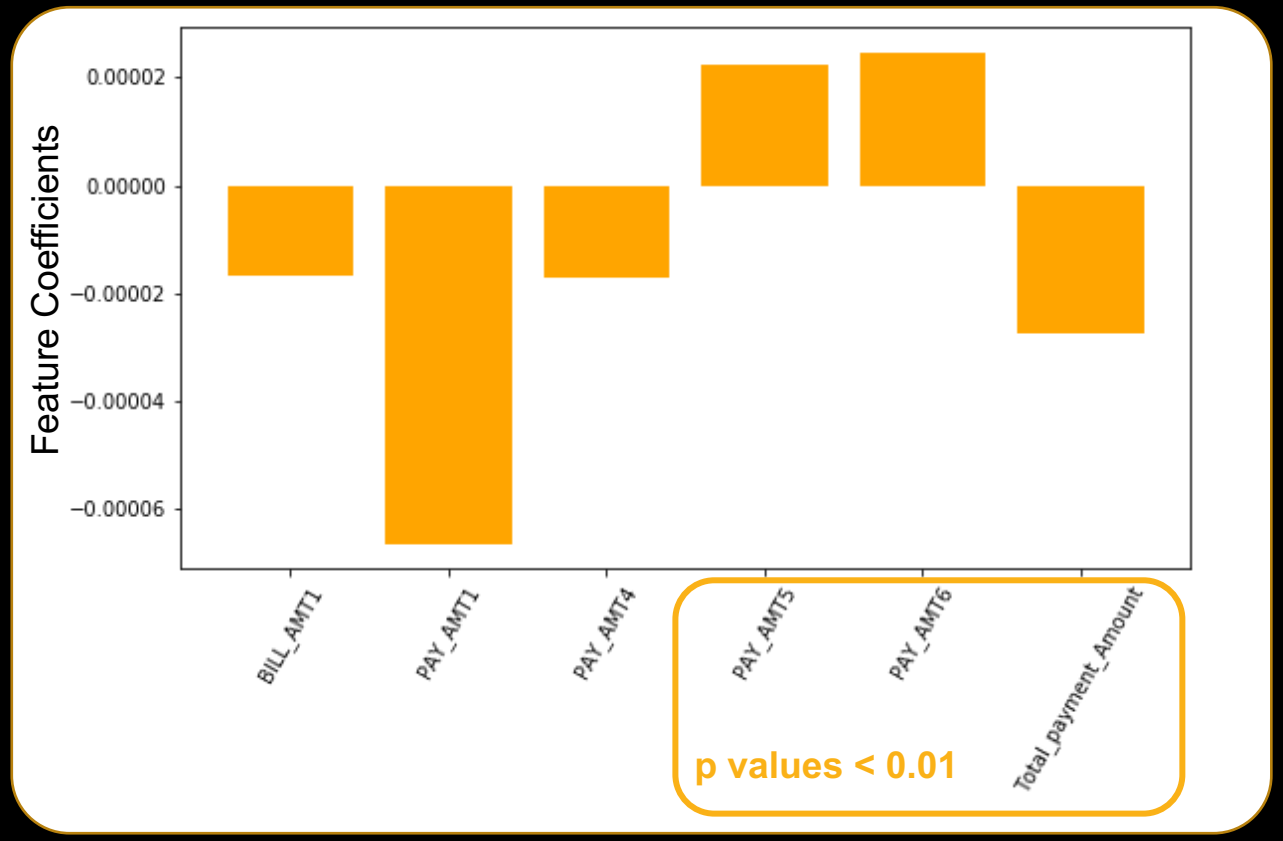
Machine Learning Interpretability Overview

A systematic workflow to drill down from global to local interpretability

Global Model Behavior

Surrogate Model – 1. Logistic Regression

A global surrogate model is an interpretable model that is trained to approximate the predictions of a black box model. We can draw conclusions about the black box model by interpreting the surrogate model.



Interpretation: Three features listed below have statistically significant effect on the probability of default but the effects are in different magnitudes and direction. For example, in the plot of feature coefficients, the increase of **total payment amount** will decrease the probability of default.

We first trained a Logistic Regression surrogate model on the original test inputs and predictions of the complex LightGBM model.

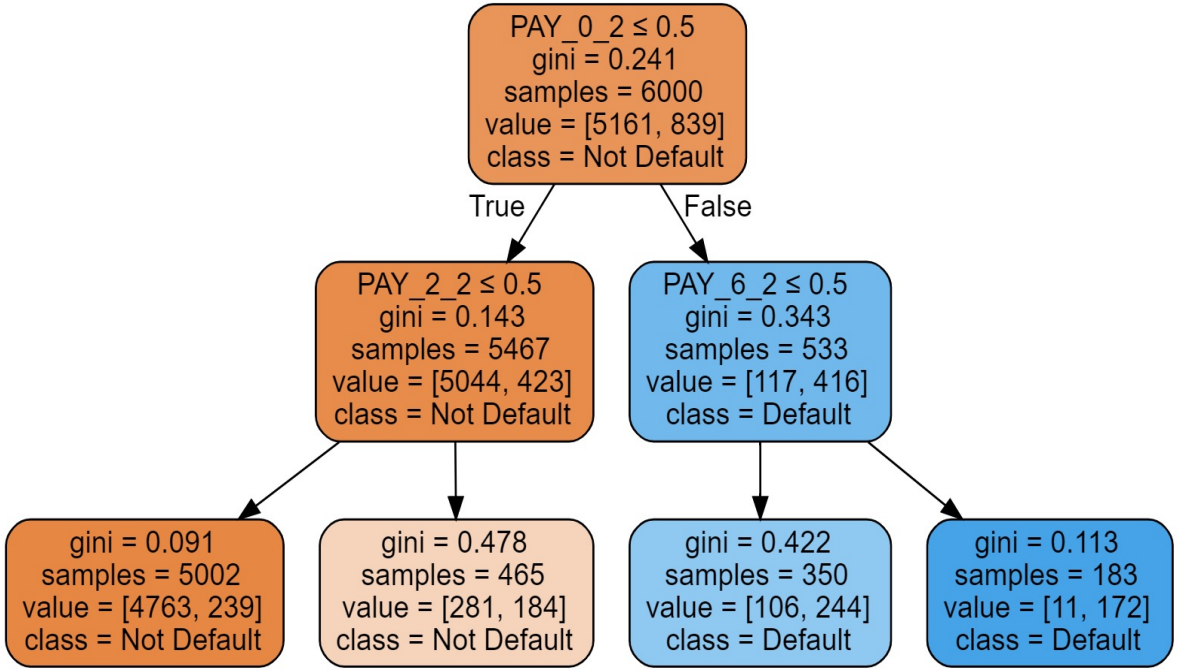
Accuracy Score: 0.78

- **PAY_AMT5**: Amount of previous payment in May, 2005 (NT dollar)
- **PAY_AMT6**: Amount of previous paymentin April, 2005 (NT dollar)
- **Total_Payment_Amount**: Summation of PAY_AMT1 – PAY_AMT6

Global Model Behavior

Surrogate Model – 2. Decision Tree

A global surrogate model is an interpretable model that is trained to approximate the predictions of a black box model. We can draw conclusions about the black box model by interpreting the surrogate model.



Interpretation: The decision tree graph approximates how the more complex LightGBM model makes predictions in an interpretable way. For example, if PAY_0=2 and PAY_6=2, then the customer is more likely to default (i.e. we move down the right-most branch of the tree).

We also trained a Decision Tree surrogate model on the original test inputs and predictions of the complex LightGBM model.

Accuracy Score: 0.811

- The variable importance, interactions, and decision paths displayed in the directed graph of the trained decision tree surrogate model are then assumed to be indicative of the internal mechanisms of the more complex LightGBM model, creating an approximate, overall flowchart for the LightGBM.

Global Feature Importance

Permutation Feature Importance

Permutation feature importance measures the importance of a feature by calculating the increase in the model's prediction error after permuting the feature. A feature is "important" if shuffling its values increases the model error, because in this case the model relied on the feature for the prediction.

Weight	Feature
0.0435 ± 0.0020	PAY_0_2
0.0202 ± 0.0035	Total_payment_Amount
0.0186 ± 0.0024	PAY_AMT6
0.0177 ± 0.0024	BILL_AMT1
0.0170 ± 0.0016	PAY_AMT3
0.0164 ± 0.0025	LIMIT_BAL
0.0145 ± 0.0023	PAY_AMT1
0.0137 ± 0.0026	PAY_AMT4
0.0133 ± 0.0036	BILL_AMT4
0.0132 ± 0.0007	BILL_AMT2
0.0120 ± 0.0017	PAY_AMT2
0.0117 ± 0.0008	BILL_AMT5
0.0112 ± 0.0025	PAY_AMT5
0.0112 ± 0.0018	Total_Bill_Amount
0.0083 ± 0.0020	PAY_0_0
0.0071 ± 0.0012	PAY_2_2
0.0064 ± 0.0018	BILL_AMT6
0.0057 ± 0.0012	BILL_AMT3
0.0041 ± 0.0015	Age_category_35 - 41
0.0035 ± 0.0015	PAY_0_-2
… 126 more …	

We get global feature importance rank for test set

The top 3 most important features in a global setting:

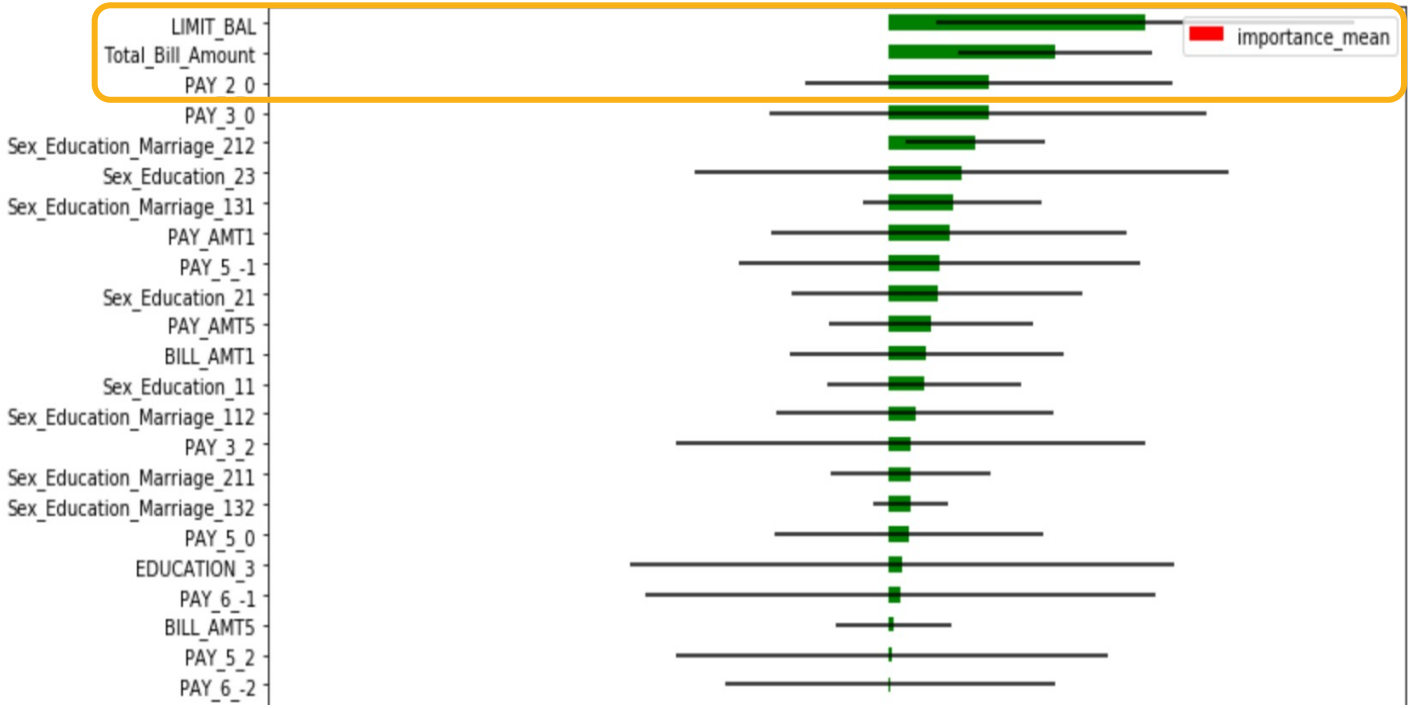
- **PAY_0_2**: Binary Variable (=1 if payment delay for two months, Repayment status in September, 2005)
- **Total_Payment_Amout**: Summation of PAY_AMT1 to PAY_AMT6
- **PAY_AMT6**: Amount of payment in April, 2005 (NT dollar)

Interpretation: If we shuffle the value of PAY_0_2, the prediction accuracy of model will experience the highest change comparing to shuffling other features. This means that, according to the PFI analysis, PAY_0_2 is the most important feature for default prediction.

Global Feature Importance

Leave One Covariate Out (LOCO)

Leave One Covariate Out measures the importance of a feature by removing one feature and calculate the error effect.



Interpretation: If we delete the feature LIMIT_BAL, the prediction accuracy of model will experience the highest change compared to when we deleted other features. This means that, for LOCO analysis, LIMIT_BAL is the most important feature for default prediction.

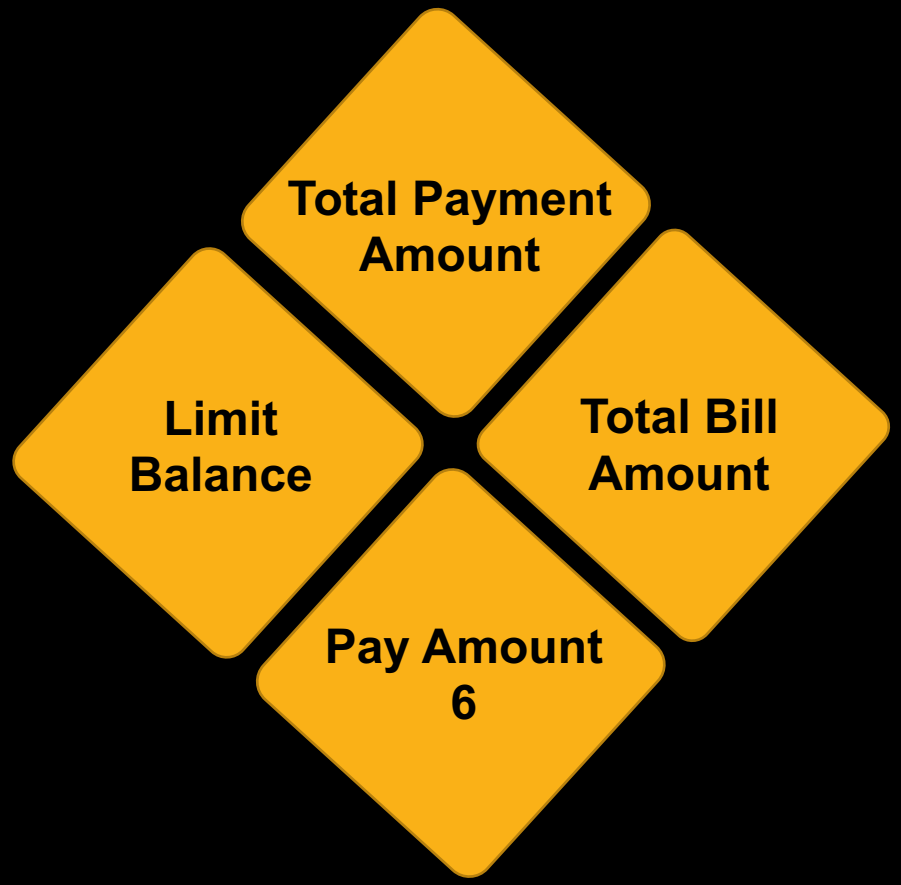
We get global feature importance rank for test set

The top 3 most important features for default prediction in a global setting:

- **LIMIT_BAL**: Amount of given credit in NT dollars (includes individual and family/supplementary credit)
- **Total_Bill_Amount**: Summation of BILL_AMT1 to BILL_AMT6
- **PAY_2_0**: Binary Variable (=1 if pay duly, Repayment status in Aug 2005)

Selected Features for Partial Dependence Plots

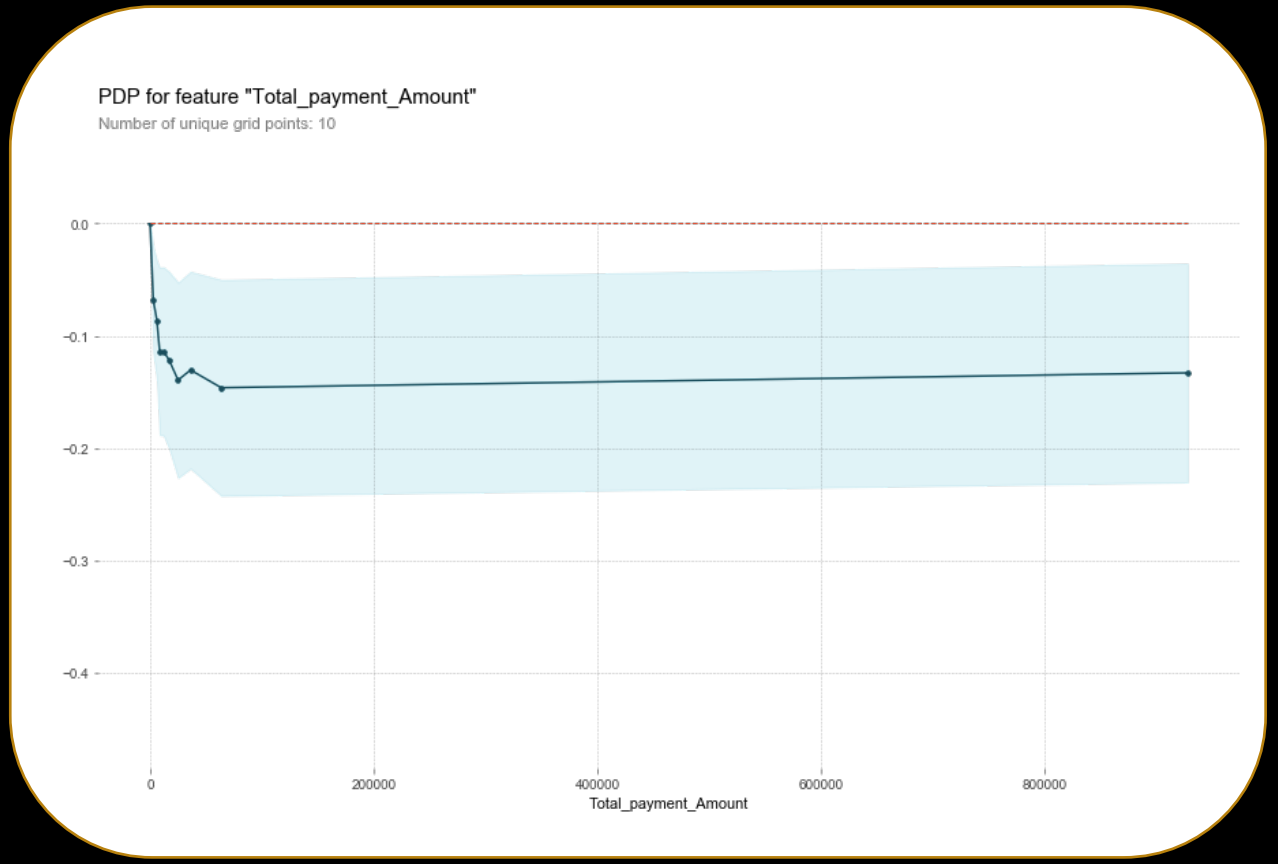
- Selected the top 4 numerical features identified by the Permutation Feature Importance and LOCO analyses for plotting of the Partial Dependence Plots



Global Feature Behavior

Partial Dependence Plot

The partial dependence plot shows the marginal effect one or two features have on the predicted outcome of a machine learning model. A partial dependence plot can show whether the relationship between the target and a feature is linear, monotonic or more complex.



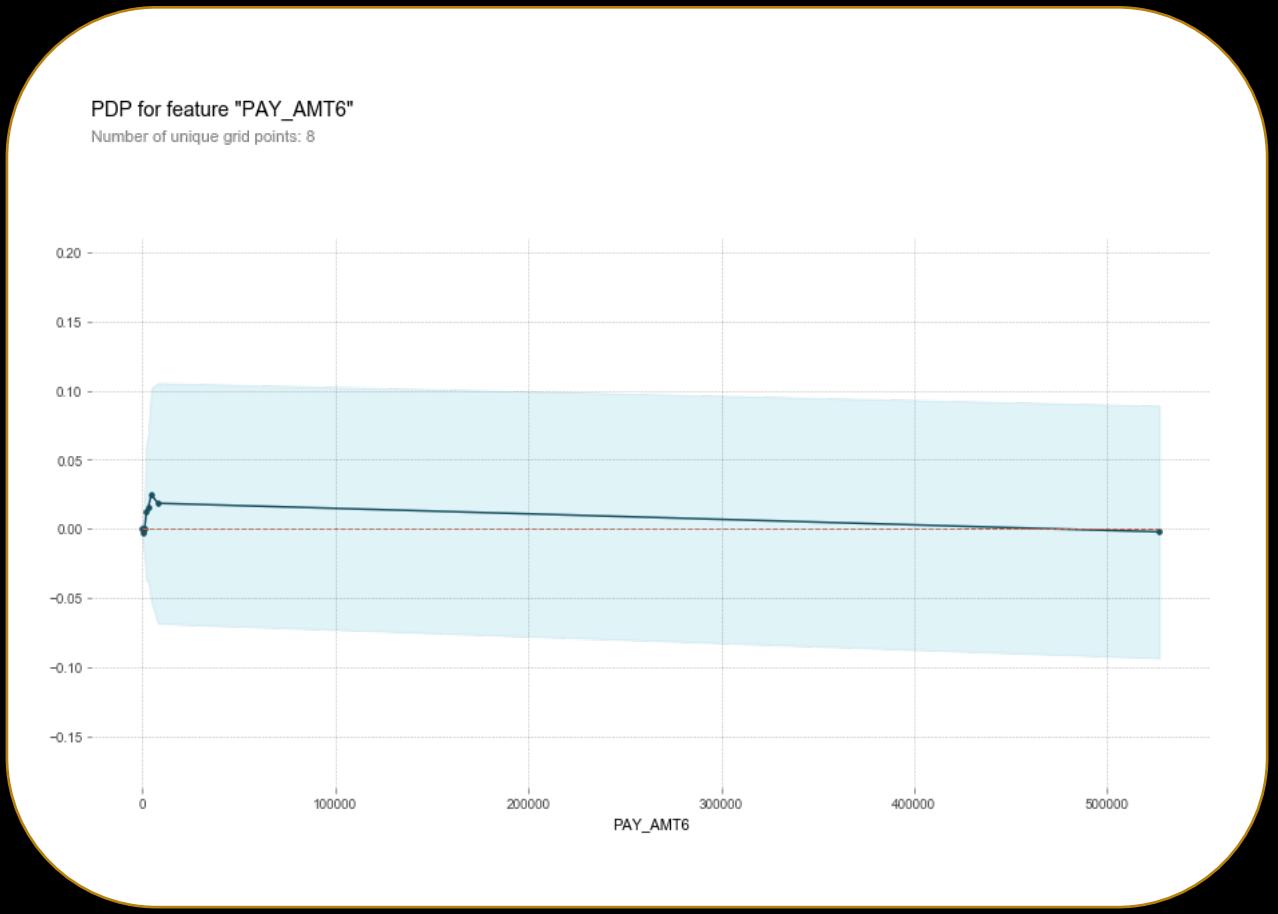
Interpretation: The marginal effect of total payment amount on probability of default shows a sharp decreasing trend when the total payment amount is less than about 100,000 NT dollars. Above 100,000 NT dollars, its marginal effect on probability of default remained almost unchanged as the total payment amount increased.

This aligns with our intuition, because someone who has been consistently paying his/her credit card bills (i.e. having a high total payment amount) is less likely to default.

Global Feature Behavior

Partial Dependence Plot

The partial dependence plot shows the marginal effect one or two features have on the predicted outcome of a machine learning model. A partial dependence plot can show whether the relationship between the target and a feature is linear, monotonic or more complex.



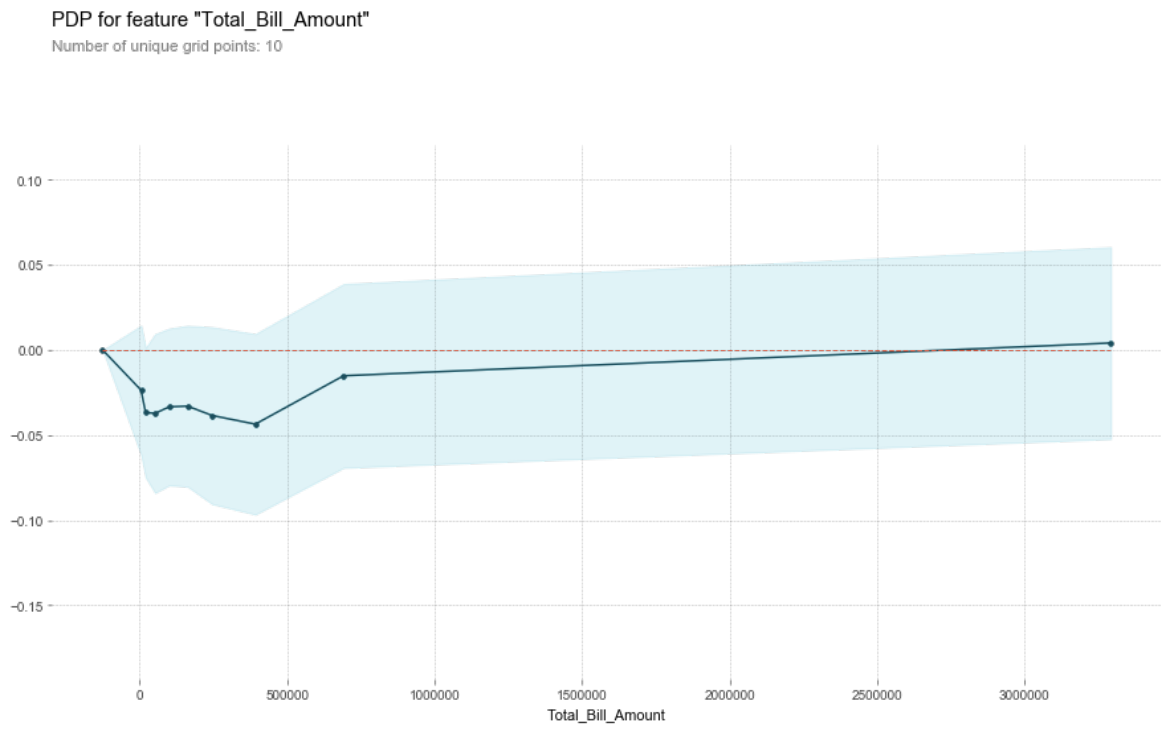
Interpretation: The **positive** marginal effect of Pay Amount 6 on probability of default is generally on an increasing trend when the amount is very small (less than about 10,000 NT dollars). However, beyond 10,000 NT dollars, its marginal effect on the probability of default decreases towards 0 as Pay Amount 6 increases.

This aligns with intuition, because, in general, the higher the greater the ability of someone to pay his/her bills (i.e. the higher the amount paid), the less likely he/she is going to default.

Global Feature Behavior

Partial Dependence Plot

The partial dependence plot shows the marginal effect one or two features have on the predicted outcome of a machine learning model. A partial dependence plot can show whether the relationship between the target and a feature is linear, monotonic or more complex.



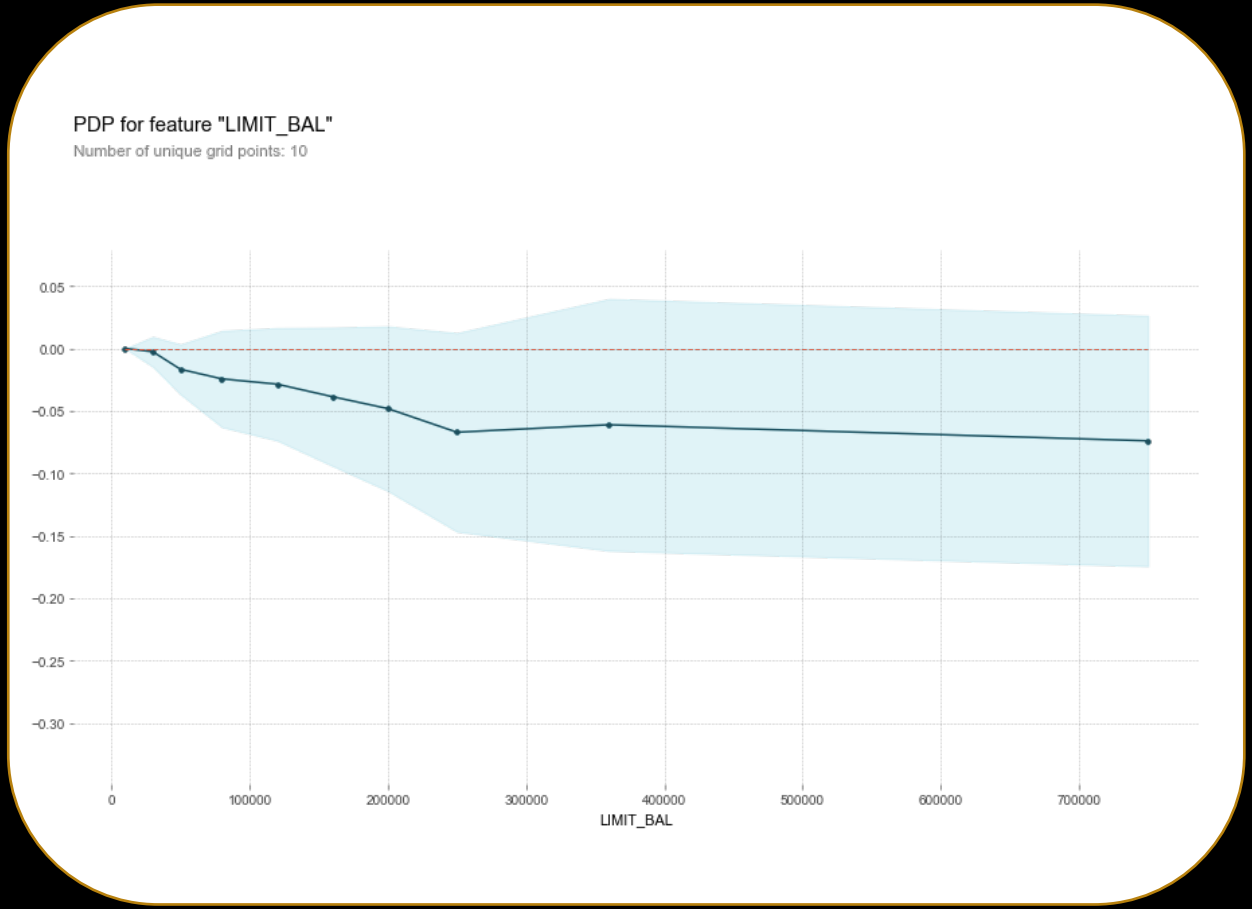
Interpretation: The marginal effect of total bill amount on probability of default first shows a decreasing trend and then an increasing trend.

This aligns with intuition, since smaller total bill amounts (i.e. clients' credit card monthly spending is consistently low) are easier to be paid in full, whereas it may be more challenging for clients to pay higher bill amounts in full.

Global Feature Behavior

Partial Dependence Plot

The partial dependence plot shows the marginal effect one or two features have on the predicted outcome of a machine learning model. A partial dependence plot can show whether the relationship between the target and a feature is linear, monotonic or more complex.



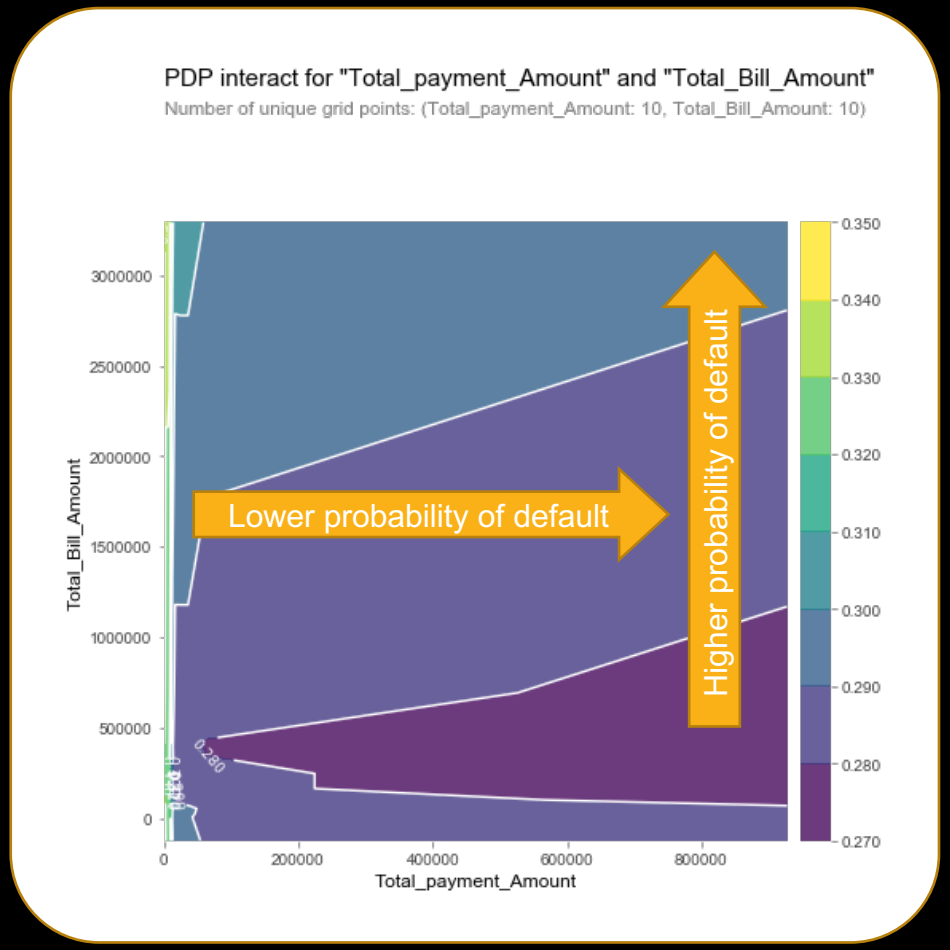
Interpretation: The marginal effect of limit balance on probability of default shows a generally decreasing trend as the limit balance increases. That is, if the customer has a higher limit balance, it is less likely for him/her to default.

The amount of given credit by the bank shows the line of credit of a customer. The higher the limit balance, the customer is deemed more reliable by the bank and is deemed more likely to pay duly.

Global Feature Behavior

2D -- Partial Dependence Plot

2D partial dependence plot is useful when we want to study the interaction between two features. Here we plot the interactions between “Total_Payment_Amount” and “Total_Bill_Amount”.



Interpretation: For a given total payment amount (i.e. when we move upwards along a vertical line on the graph, especially when the total payment amount is low), the default probability generally increases when a client clocks higher bills.

For a given total bill amount, the probability of default generally decreases as we move rightwards along a horizontal line on the graph.

These observations aligns with intuition, since someone is more likely to default if he/she clocks higher bills. Also, someone is less likely to default if he/she is able to pay for his/her credit card bill in higher amounts.

Local Feature Importance

LIME

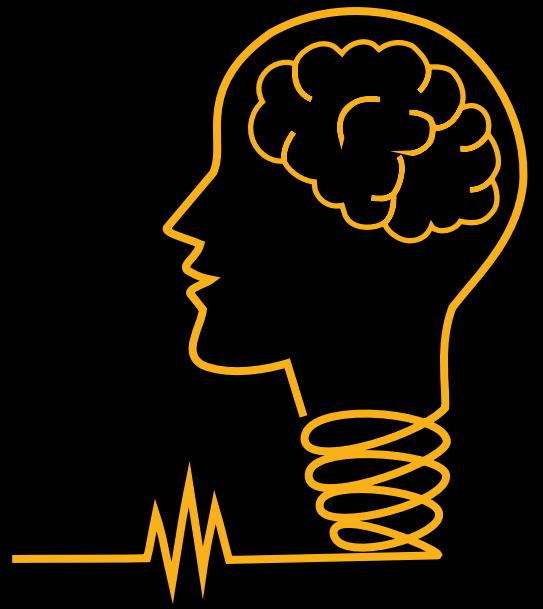
An interpretable surrogate model used to explain non-linear decision boundary in local region.

Shapley Value

The value of feature contributed to the prediction of the particular instance compared to average prediction for the dataset.

Instances

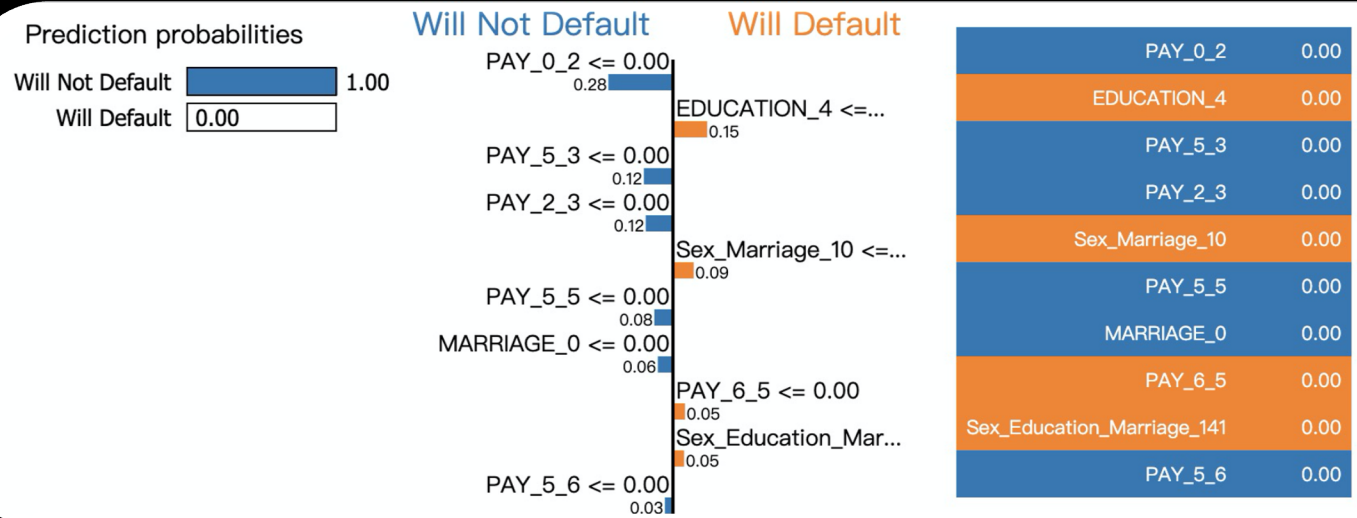
Here, to illustrate the interpretation of LIME and Shapley Values over a range of default probabilities, we use 0 percentile, 50 percentile and 99 percentile as instances to visualise the LIME and Shapley Value results.



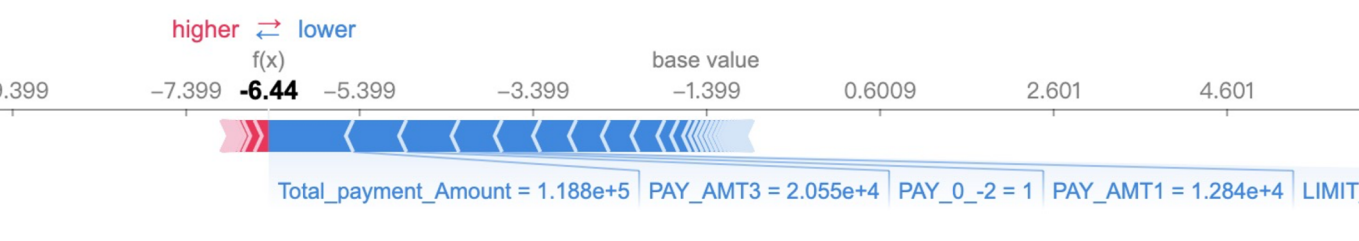
Local Feature Importance

Instance 1: 0th percentile sample

LIME



Shap force plot



Sigmoid Transformation of –6.44: 0.001 probability of default

Interpretation:

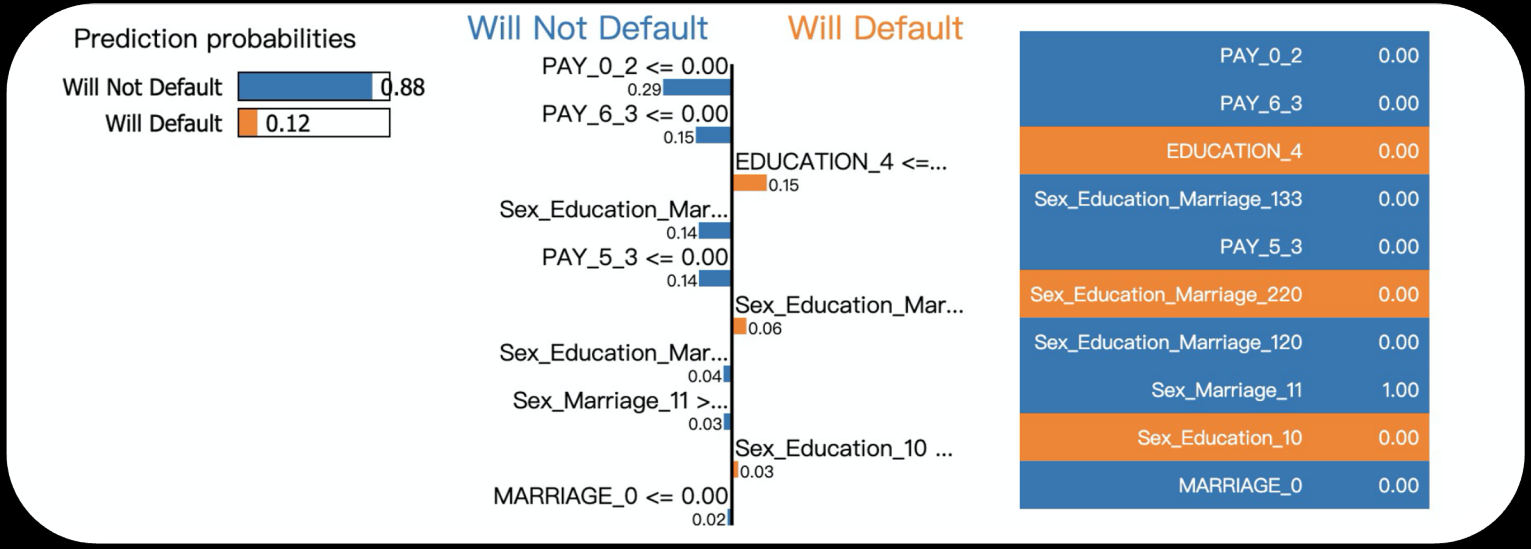
- LIME - Model predicts that the client will not default. The strongest effect is the fact that the client's payment was not delayed for 2 months in Sep 2005 (i.e. the condition Pay_0 = 2 is false).
- Shap - The biggest impact on the prediction of non-default comes from the fact that total payment amount equals to 118,800 NT dollars (significantly higher than the mean of 31,651 NT dollars*) which has significantly decreased the probability of default.

* Calculated based on total payment amounts as captured in the original dataset

Local Feature Importance

Instance 2: 50th percentile sample

LIME



Shap force plot

Sigmoid Transformation of –1.95: 0.12 probability of default

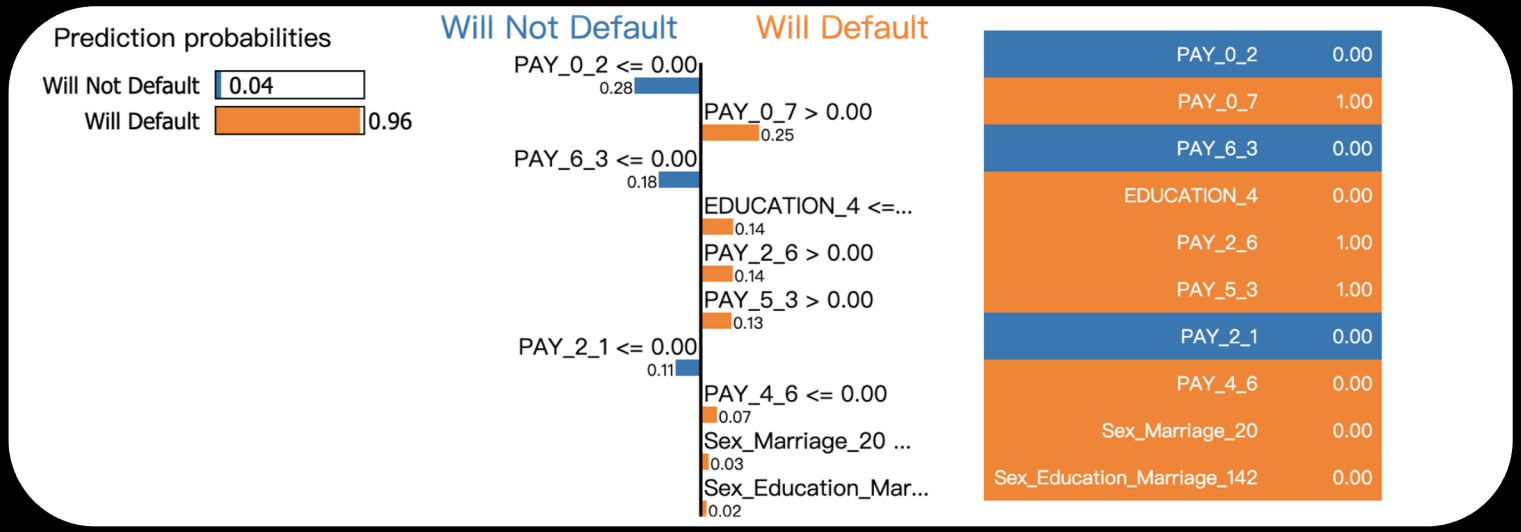
Interpretation:

- LIME - Model predicted that the client is highly unlikely to default. Biggest effect is also the fact that the client's payment was not delayed for 2 months in Sep 2005.
- Shap - The biggest impact comes from the fact that the amount of payment in Jul 2005 equals to 10,020 NT dollars which will decrease the probability of default, although the repayment status in May 2005 (delayed for 2 months) has a significant effect on increasing the probability of default.

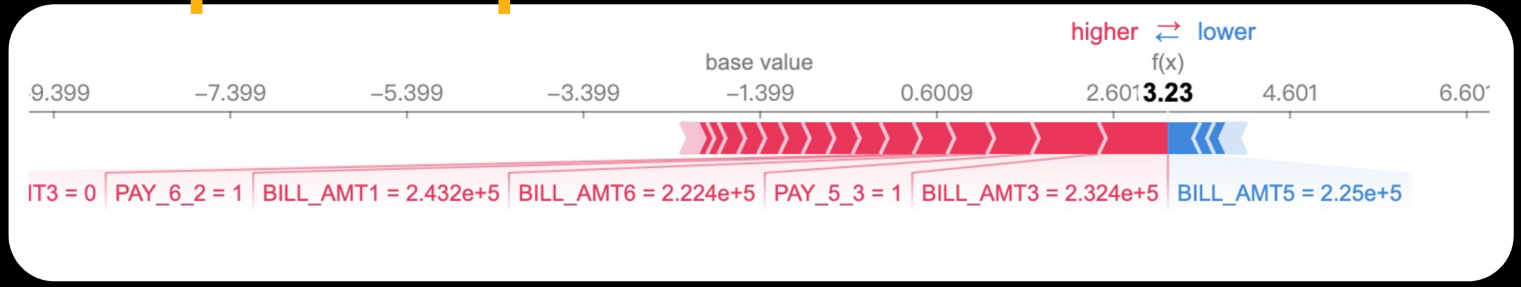
Local Feature Importance

Instance 3: 99th percentile sample

LIME



Shap force plot



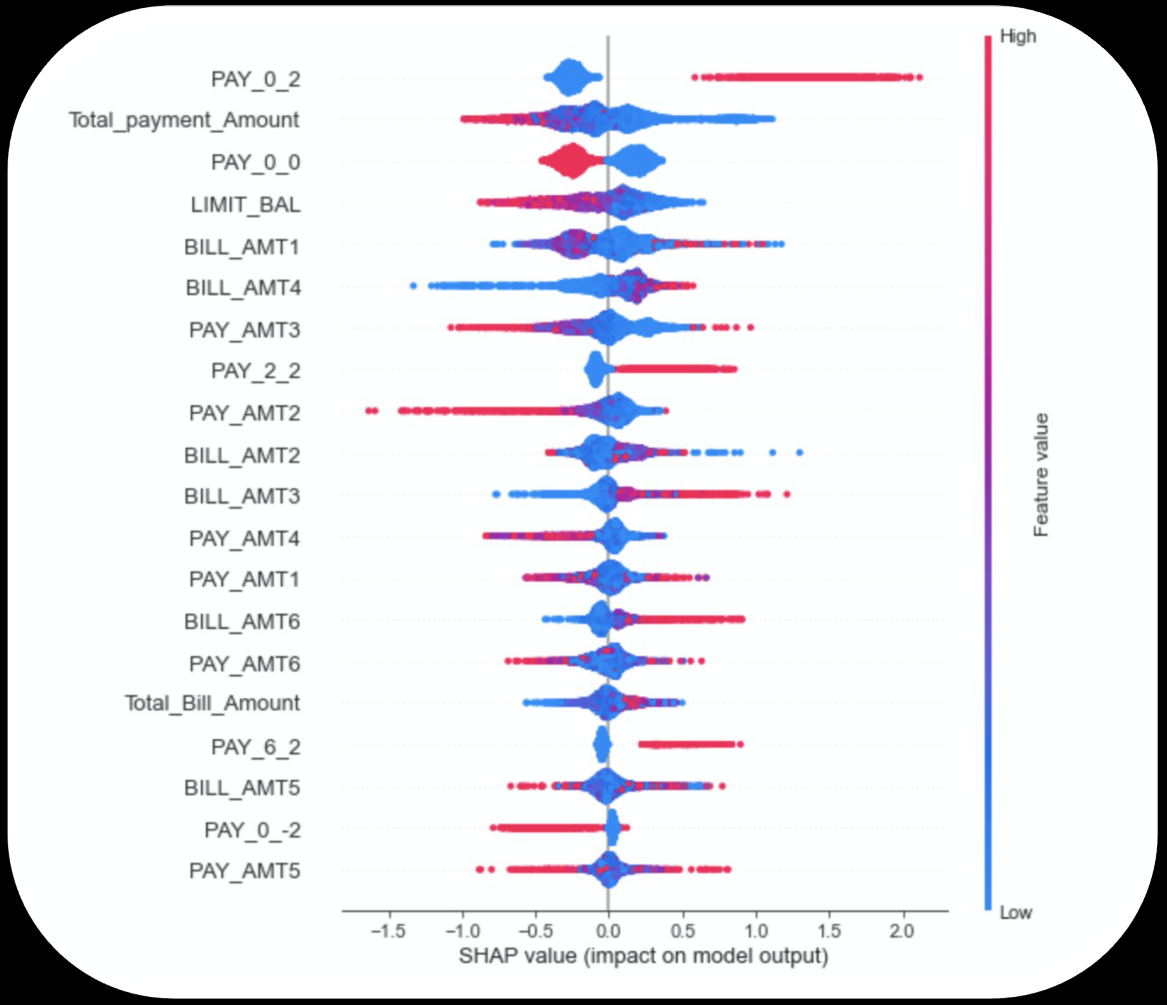
Sigmoid Transformation of 3.23: 0.96 probability of default

Interpretation:

- LIME - Model predict that the client will default. The biggest contributor of likelihood of default is the fact that the client's payment was delayed for 7 months in Sep 2005.
- Shap - The biggest impact comes from the fact that the client's bill amount in Jul 2005 amounted to 232,400 NT dollars which will increase the probability of default. Another significant contributing factor was the fact that the client's repayment status was delayed for 3 months in May 2005.

Local Feature Importance

Summary Plot For Shapley Value



Interpretation:

- The distinct separation of the red and blue dots for PAY_0_2 suggests that the client's 2-month delay of payment in Sep 2005 will certainly increase his/her likelihood of default in Oct 2005.
  - The wide spread of the pink dots for PAY_0_2 also suggests that this feature had varying effects on the probability of default for different clients

- Based on the plot of the 2nd most impactful feature, the total payment amount, in general, a client is less likely to default when the total payment amount is high and vice-versa
  - However, this trend is less generalisable compared to that indicated by PAY_0_2 since the red and blue dots are not as clearly separated

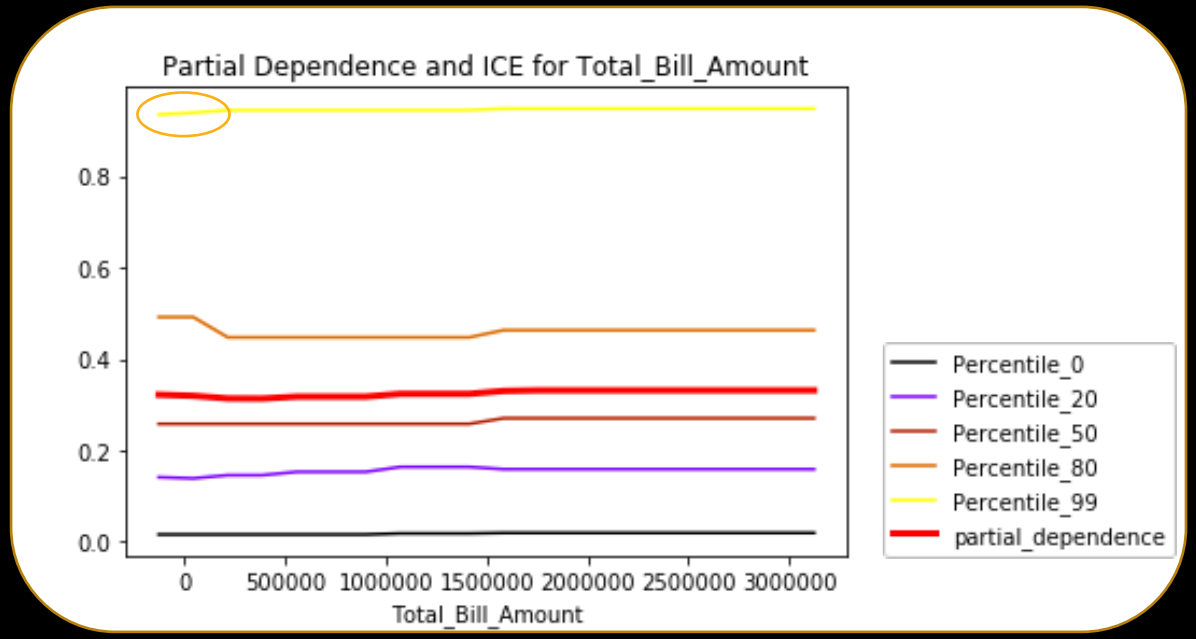
Local Feature Behavior

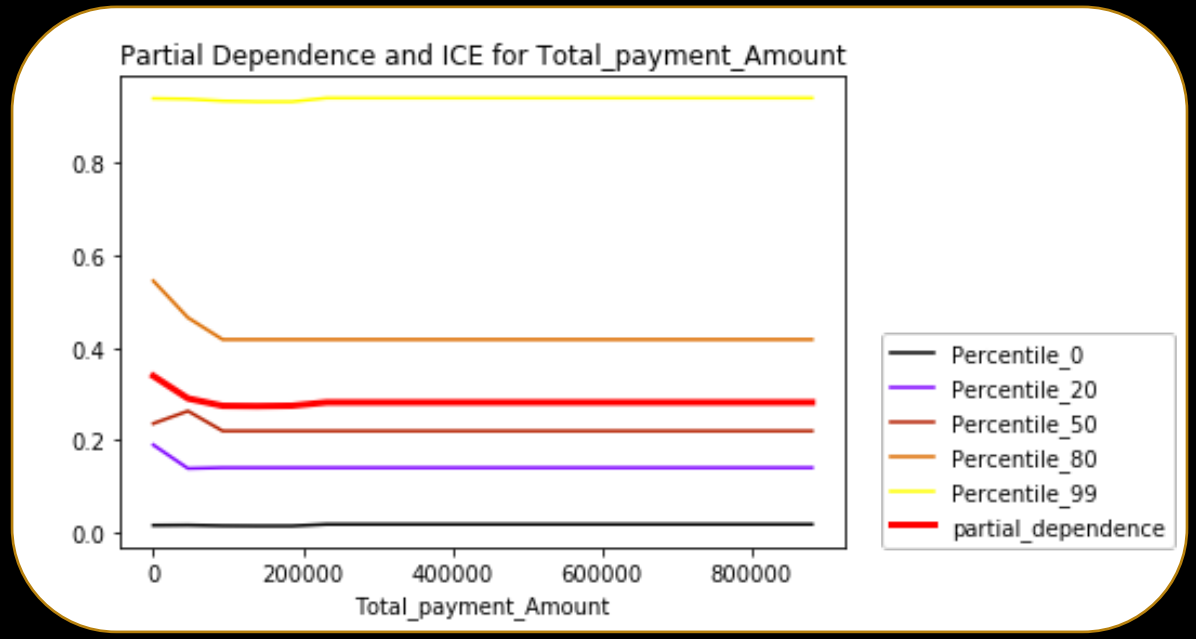
Individual Conditional Expectations

- A local interpretability measure
- Visualize the dependence of the prediction on a feature for each instance separately
- Visualized instances from the 0^{th}, 20^{th}, 50^{th}, 80^{th} and 99^{th} percentile

Interpretation:

- Total bill amount: The ICE plot for total bill amount suggests that the probability of default is generally unchanged as the total bill amount increased. For those people who are likely to default (see yellow line encircled), the probability of default increases slightly as the total bill amount increases from 0 to 100,000 NT dollars.
- Total payment amount: On an aggregated level (i.e. see line in red), the probability of default decreases slightly as total payment amount increased from 0 to about 200,000 NT dollars. Beyond 200,000 NT dollars, the effect on probability of default is unchanged.





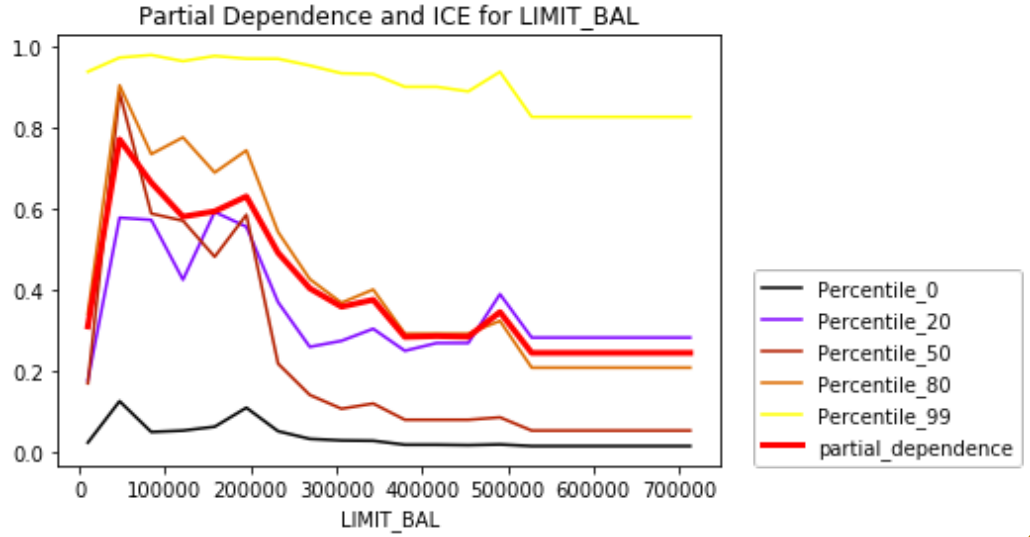
Local Feature Behavior

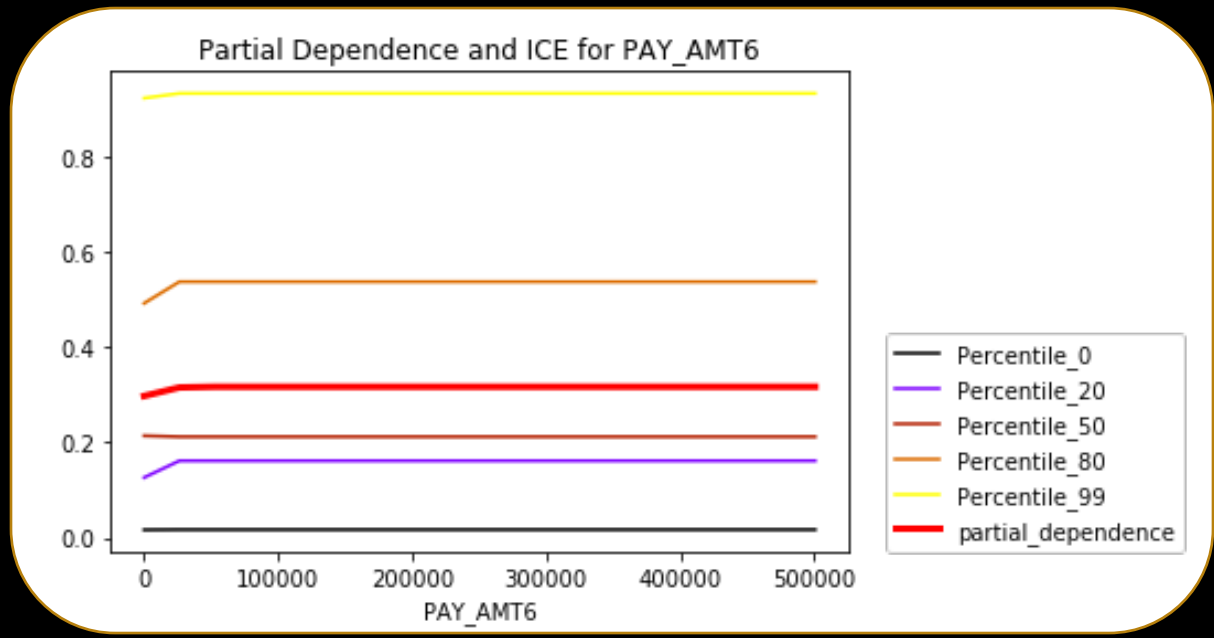
Individual Conditional Expectations

- A local interpretability measure
- Visualize the dependence of the prediction on a feature for each instance separately
- Visualized instances from the 0^{th}, 20^{th}, 50^{th}, 80^{th} and 99^{th} percentile

Interpretation:

- Limit balance: For the range of limit balance [0, 50,000], the default probability increases sharply with increasing limit balance. Comparatively, it is not as clear (i.e. fluctuating) in the range [50,000, 200,000], although the probability of default generally decreases with increase in limit balance
- Pay amount 6: Within the range of pay amount 6 [0, 25,000], the probability of default generally increases slightly with increasing payment amounts. Beyond that, the probability of default is unaffected by the payment amount.





Observations

1. Different MLI techniques may identify different features which are important for default predictions
   - PFI's important features – PAY_0_2, Total_Payment_Amount and PAY_AMT6
   - LOCO's important features – LIMIT_BAL, Total_Bill_Amount and PAY_2_0

   Domain knowledge is still helpful or even essential for the interpretation of results
2. Clients' behaviour in the months Apr and Sep 2005 were especially insightful for default probability predictions
   - Amount of payment in **Apr 2005 –** PAY_AMT6
     - Statistically significant feature in according to the Logistic Regression surrogate model
     - Important feature in the Permutation Feature Importance table
   - Repayment status in **Sep 2005** – PAY_0
     - Important feature as identified by the Decision Tree surrogate model
     - Important feature in the Permutation Feature Importance table and Shapley Value summary plot
3. Local Feature Behaviour plots can help to effectively explain default probability trends for different clients as we vary the magnitude and direction of the identified feature
   - Helps to provide business insights for treating different segments of customers

Model Limitations

1. Recall score could be improved if we had more data for model training

2. Lack of macro data such as:
   - Performance of economy
   - Clients' behaviour score
   - Credit card ownership from other banks

3. Lack of clients' microdata such as:
   - Occupation
   - Income

Such features can further *enrich* our dataset for better prediction performance and explainability

Thank You