

## Reporte Modelaje One Credit.

### Descripción del Problema.

Se ha presentado un incremento en las tasas de no pago o “default” que han presentado los clientes recomendados por la empresa aprobadora de créditos “Credit One”.

Para ello se ha analizado la información de los últimos seis meses de un compendio de registro de pagos de 30000 clientes para analizar que atributos o conjunto de los registros funcionan como indicadores clave para conocer si un cliente entrará en no pago o “default”.

### Características de la Información Analizada. Principales Hallazgos

Al analizar la información presentada se tienen las siguientes características.

SEX EDUCATION		Mujer				Hombre				Total	
		graduate school	high school	others	university	graduate school	high school	others	university		
Default proximo mes	Rango_edad										
Pago regular	21_27	1273	302	62	1767	533	192	32	845	5006	
Pago regular	28_30	1084	181	34	910	615	113	23	576	3536	
Pago regular	31_37	1536	431	77	1843	1073	263	36	1077	6336	
Pago regular	38_43	675	460	44	1161	592	292	26	706	3956	
Pago regular	44_79	533	861	62	1053	635	585	39	762	4530	23364
No pago o “default”	21_27	319	108	5	610	132	86	1	337	1598	
No pago o “default”	28_30	218	48	2	248	130	51	1	175	873	
No pago o “default”	31_37	314	109	3	453	286	100	2	342	1609	
No pago o “default”	38_43	152	136	4	285	172	95	3	253	1100	
No pago o “default”	44_79	127	291	5	326	186	213	7	301	1456	6636
		6231	2927	298	8656	4354	1990	170	5374	30000	

Tabla #1: Resumen características sociales de los clientes con respecto a su estado de pago.

Del total de registros 23364 corresponden a registros que están dentro de pago regular y 6636 en “default”. Del total de registros 18112, son de mujeres 74% están en pago regular y 26% en no pago y 11888 de hombres, 68% están en pago regular y 32% en no pago.

Por rango de Edad y Educación se tiene lo mostrado en las tablas #2 y #3

SEX	Mujer				Hombre			
EDUCATION	graduate school	high school	others	university	graduate school	high school	others	university
Pago regular	5101	2235	279	6734	3448	1445	156	3966
No pago o “default”	1130	692	19	1922	906	545	14	1408
Total	6231	2927	298	8656	4354	1990	170	5374

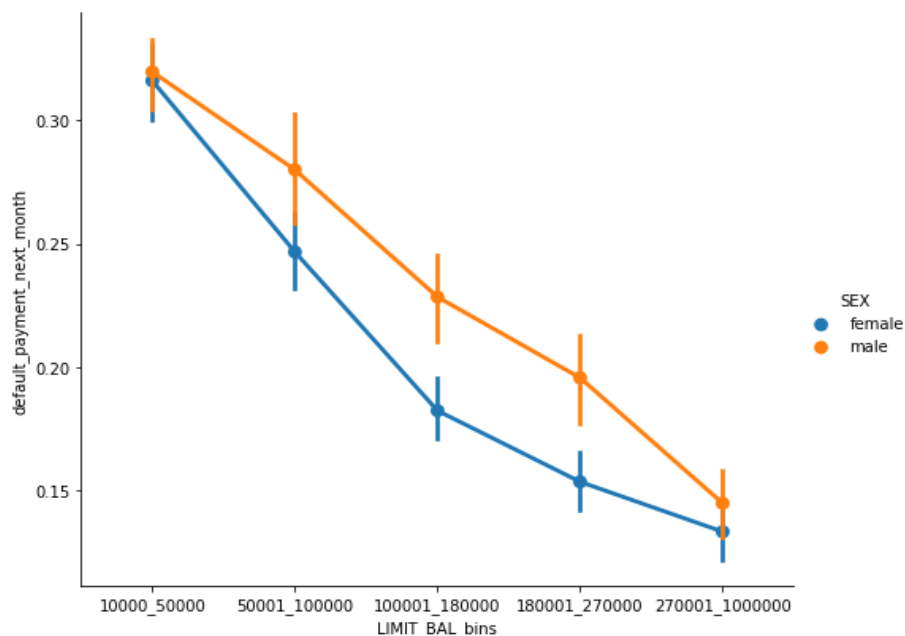
Tabla #2: Resumen características sociales de los clientes con respecto a género y grado educativo.

SEX		Total
EDUCATION		
Default proximo mes	Rango edad	
Pago regular	21_27	5006
Pago regular	28_30	3536
Pago regular	31_37	6336
Pago regular	38_43	3956
Pago regular	44_79	4530
No pago o "default"	21_27	1598
No pago o "default"	28_30	873
No pago o "default"	31_37	1609
No pago o "default"	38_43	1100
No pago o "default"	44_79	1456

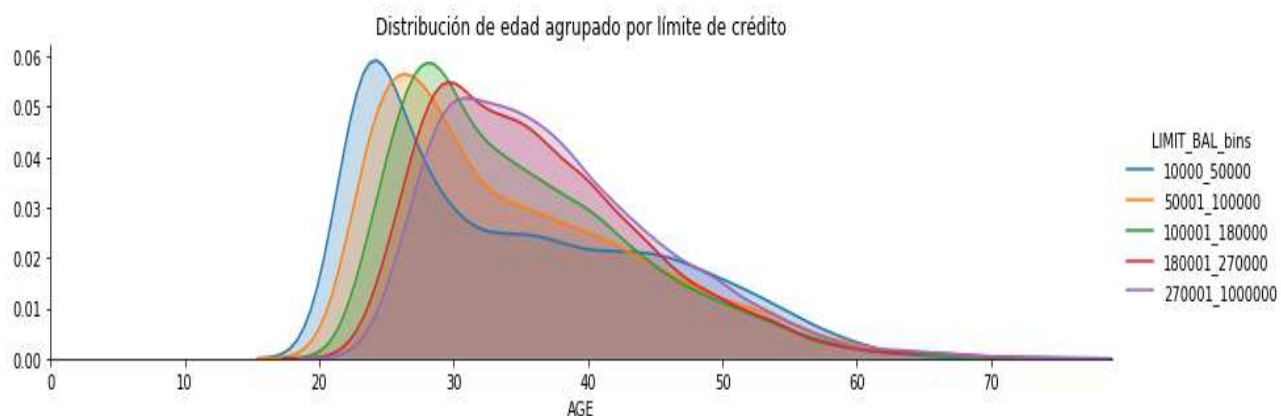
**Tabla #2: Resumen características sociales de los clientes con respecto a edad.**

Del estudio detallado, de las tablas 2 y 3 se denota que no existe una diferencia significativa entre los estado de no pago y pago regular entre los distintos grupos de género, grado educativo y edad.

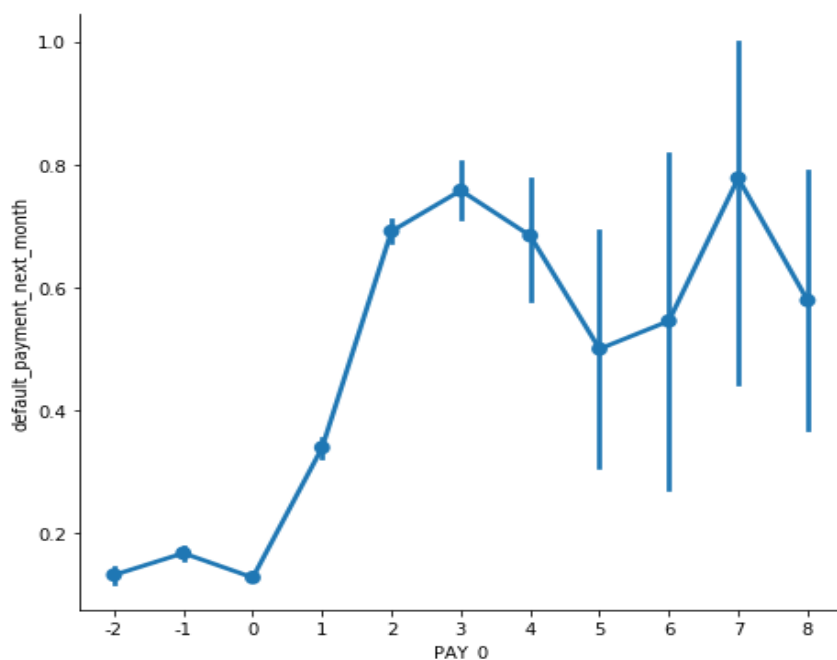
Ahora bien analizando los grupos de límite de crédito en función del estado de pago se tiene lo que se visualiza en la gráfica #1. Se nota que existe una diferencia signitificativa en en cuanto a cada rango de límite de crédito, siendo el rango de crédito de 10000-50000 el que tiene una probabilidad del 0,3% de entrar en no pago. Casi el doble del último límite de crédito de 0,15%.



Las características más importantes de comportamiento se nota al dividir el atributo de límite de crédito contra la posibilidad de entrar en default, el rango de 10000-50000 tiene una mayor probabilidad de entrar en default, independientemente del género, grado educativo y edad. Aunque si se logra apreciar que este grupo de crédito es el que tiene una mayor cantidad de usuarios dentro del rango de 21-27 años, aunque no es exclusivo del mismo.



Otro indicador importante que resalta a simple vista es como los estados de pago -2,-1 y 0 en cualquiera de los 6 meses anteriores presenta una probabilidad baja de entrar en impago. Estos representan los clientes. Si los estados de pago entran en demora mayor a 1 mes, la probabilidad y la variabilidad de entrar en default aumenta.



### **Modelos Probados.**

Para realizar el modelado clasificatorio se procedió a codificar de manera adicional los atributos existentes, asignando valores numéricos a cada categorías de los atributos Edad, Limite Crédito, Pago actual xx o “PAYxx”, donde xx representa cada mes anterior, Bill\_ATMxx y PAY\_ATMxxx. Similarmente se codificaron las categorías EDUCATION y MARRIAGE o Estado civil con “one shot encoding”. Así los atributos finales para modelado y optimización son:

SEX
PAY_0
PAY_2
PAY_3
PAY_4
PAY_5
PAY_6
LIMIT_BAL_bins
LIMIT_AGE_bins
BILL_ATM1_bins
BILL_ATM2_bins
BILL_ATM3_bins
BILL_ATM4_bins
BILL_ATM5_bins
BILL_ATM6_bins

PAY_AMT1_bins
PAY_AMT2_bins
PAY_AMT3_bins
PAY_AMT4_bins
PAY_AMT5_bins
PAY_AMT6_bins
EDUCATION_high school
EDUCATION_others
EDUCATION_university
MARRIAGE_married
MARRIAGE_others
MARRIAGE_single

**Tabla 4 y 5: Atributos utilizados para los modelos.**

Los resultados obtenidos de los tres modelos escogidos se presentan a continuación:

	R Squared	RMSE
SVM	-0,061	<b>0,426</b>
RF	-0,152	<b>0,444</b>
KNN	-0,201	<b>0,453</b>

**Tabla 6: Resultados de los modelos SVM,RF y KNN. Con datos de entrenamiento.**

### **Resultados del Modelo Escogido.**

Una vez realizado, los modelos con datos de entrenamiento se procede a predecir los resultados con los datos de prueba. Comparando los valores predecidos con los reales.

Los resultados obtenidos, se obtienen los siguientes resultados en Precisión y AUC. (“Area Under the Curve”), área debajo de la curva.

Modelo	Precisión	AUC	VN	FP
SVM	0,82	0,73	0,31	0,17
Random Forest	0,77	0,77	0,51	0,13
KNN	0,81	0,76	0,34	0,18

**Tabla 6: Resultados de los modelos SVM,RF y KNN. Con datos de entrenamiento.**

Los modelos que se comportan mejor son los modelos de Random Forest y KN, debido al valor AUC. Sin embargo también se debe analizar los valores de verdaderos negativos (VN) de la tablas de confusión, esto es valores que son “1”, es decir cliente en default, o no pago que el modelo predice como de pago regular o “0”.

Como se puede observar los modelos SVM y KNN tienen la tasa más baja de esta condición. Esto es la conidición más delicada puesto el modelo pasaría por alto o predeciría equivocadamente un cliente como pago regular cuando debería ser en default. Por ello los modelos a utilizar deberían ser SVM o KNN. Esto a pesar de que en estos casos los modelos arrojarían más casos de falsos positivos, es decir clientes que son de pago regular o “no default” que el modelo predice como de default.

Bien se debería valor el costo de recurso que se perdería por evaluar un caso que no entrará en default. Si el costo de tratar un cliente no default como default es muy alto bien se puede utilizar el modelo Random Forest.

Además de los modelos obtenidos se utilizó el modelo de Random Forest para deducir las reglas más importantes que se pueden obtener.

#### **No Default**

`PAY_0 <= 4.5 and PAY_AMT3_bins > 1.5 and PAY_2 <= 4.5 and PAY_5 <= 4.0`

Esta regla indicar si el estado del mes actual , anterior, como el mes 5 es cualquiera de -2,-1,0,1, desde no utlizar el crédito hasta 1 mes de atraso. Además si el pago del mes trasanterior es mayor a \$3. Existe un a precisión del 88% de los casos donde no entrará en default.

### **Default**

PAY\_0 > 4.5 and PAY\_2 <= 5.5 and PAY\_2 > 2.5 and EDUCATION\_others <= 0.5

Esta regla obedece a las siguiente condicion:

Estado de Pago Mes actual es mayor a estado 1(1 mes de atraso), el estado del mes anterior es 0,1 o 2 (rango de pagó a tiempo a 2 meses de atraso) y grado educativo no es "otros". Existe una precisión del 72% de los casos donde entrará en default.

### **Conclusiones y Recomendaciones Adicionales.**

Como se podría observar existe una dependencia alta del estado del mes actual PAY\_0 en determinar si el cliente entra en default con independencia de los demás atributos. Por lo tanto se podría determinar en futuros modelos que otros atributos pueden predecir las categorías de este atributo, utilizandolo como valor meta o "target" a predecir.

Las características anteriormente descritas por si solas permiten al negocio formular políticas para formentar el pago a tiempo de los créditos o reformular las políticas al brindar limites de crédito. O bien reducir la cartera de clientes donde se concentren bajos límites de crédito.

Bien del análisis global realizado, se realizan las siguiente recomendaciones:

- 1- Aumentar la cartera de clientes con mayor límite de crédito pues tienen a tener mejor respuesta a pagar a tiempo.
- 2- Fomentar políticas para el pago a tiempo para la cartera de clientes con límite de crédito de 10000-50000, donde la misma está constituida mayoritariamente por personas del rango de edad de 21-27 años.
- 3- Fomentar políticas para que los usuarios, cualquiera que sea su condición educativa, de género y edad estén al día con su pago. Pues del análisis exploratorio se nota como es posible que conforme se aumente los meses de pago en mora aunque sea 1 mes, aumenta la posibilidad de impago.
- 4- Realizar un modelado de datos donde se permita visibilizar mejor que combinación especifica de edad, educación y genero, balance de crédito permite identificar nichos más especificos de mercados donde se les pueda aplicar políticas orientadas a los mismos.
- 5-Detectar el comportamiento riesgoso de los clientes utilizando los modelos y reglas aquí propuestas, y así enfocar las políticas de control de riesgo previamente detallas en los casos que lo predigan.