

# Predicción de riesgo crediticio: contexto de la crisis bancaria de 2023

Entrega final CoderHouse  
Alejandro Senabre



# Agenda



## 1. Introducción

Se dará un contexto para entender la problemática a nivel mundial. Se usará como ejemplo la quiebra del Silicon Valley Bank.



## 2. Metadata

Breve análisis y características del conjunto de datos extraído para la realización de este modelo.



## 3. Análisis Gráfico/Estadístico

Se trabajará a partir de gráficos para entender el impacto de la tasa de interés en la morosidad.



## 4. Modelización y Conclusión

Se hará un modelo que sea capaz de predecir la morosidad de un préstamo.

# 1.

# Introducción

## Contexto:

En marzo de 2023, el sistema bancario mundial sufrió un revés significativo cuando el Silvergate Bank y el Silicon Valley Bank, uno de los bancos más importantes de Estados Unidos y principal inversor en empresas tecnológicas emergentes, se declararon en quiebra. Este evento marcó la segunda falla bancaria más grande en la historia de Estados Unidos, solo superada por el colapso de Lehman Brothers Holdings durante la crisis financiera de 2008.

En este trabajo se intentará mostrar uno de los principales responsables del problema que tienen los bancos y tratar de predecir, a partir de datos extraídos de préstamos bancarios, la probabilidad que los préstamos dados por el banco terminen convirtiéndose en morosos.

## Preguntas de interés:

¿Cómo puede llegar a afectar la tasa de interés a los préstamos dados?

¿Hay forma de verificar que tan probable es que un préstamo termine en estado de deudor?

Ante una subida de la tasa de interés, ¿Cómo afectará a los préstamos ya dados?

1.

# Introducción

## Silicon Valley Bank



Silicon Valley Bank (SVB) fue un banco estadounidense fundado en 1983 en Santa Clara, California. SVB se ha centrado en atender a empresas de tecnología asentadas en la costa oeste de EE.UU. y ha sido un prestamista crucial para muchas compañías emergentes.

En la última década, SVB experimentó una rápida expansión. El banco era conocido por trabajar con nuevas empresas tecnológicas jóvenes que otros bancos pueden haber evitado. Cuando esas nuevas empresas florecieron, el SVB creció junto con ellas.

Pero en el último año, debido a las constantes subas de la tasa de interés en Estados Unidos para hacer frente a la inflación que está sufriendo el país, llevó al banco a una falta de liquidez para hacer frente a sus demandas, provocando el pánico de sus clientes y el posterior desplome del banco.



# 1.

## Introducción

Aunque hubo múltiples razones mas que llevaron a la quiebra del SVB, el **aumento de la tasa de interés** es un factor que puede perjudicar gravemente a los bancos, esto lo podemos observar con las siguientes noticias:

### *Big Banks Set Aside Billions as They Brace for a Downturn*

The country's largest lenders increased their reserves to protect against deteriorating economic conditions this year, after reporting resilient profits for the end of last year.

*Stacy Cowley & Rob Copeland, The New York Times, Jan 13, 2023*

Esta noticia extraída del *The New York Times* nos habla de como los principales bancos de Estados Unidos se están preparando para una posible recesión en el futuro cercano. Están aumentando sus reservas y pronostican un continuo aumento de las tasas de interés por la continua escala de la inflación que sufre el país. Esto puede generar problemas de morosidad y liquidez, lo que aumenta el riesgo de no poder devolver el dinero depositado por los ahorradores y podría llevar al colapso del banco.

## 2.

# Metadata

### Data:

El dataset fue obtenido a través de la página web de [Kaggle](#), de datos que son parte del IIIT Bangalore Case Study, Instituto Internacional de Tecnología de la Información de Bangalore, una universidad de investigación de India.

El origen de los datos no está especificado, por lo que no se sabe si es un dataset de un banco real o datos artificiales creados para este instituto, pero de todas formas nos van a resultar útiles para el análisis.

Se hará una breve descripción de los datos:

Variable	Descripción
loan_amnt	La cantidad de dinero solicitada por el prestatario para el préstamo.
term	El número de pagos en el préstamo, expresado en meses, que puede ser 36 o 60.
int_rate	La tasa de interés del préstamo.
installment	El monto del pago mensual que el prestatario debe pagar si se origina el préstamo.
grade	La calificación asignada al préstamo por LC (LendingClub) en función de la solvencia crediticia.
sub_grade	Una subcalificación más detallada asignada al préstamo por LC.
emp_title	El título laboral proporcionado por el prestatario al solicitar el préstamo.
emp_length	La antigüedad laboral en años. Los valores van de 0 (menos de un año) a 10 (diez o más años).
home_ownership	El estado de propiedad de la vivienda proporcionado por el prestatario durante el registro.
annual_inc	Los ingresos anuales autodeclarados proporcionados por el prestatario durante el registro.
verification_status	Indica si los ingresos fueron verificados por LC, no verificados o si la fuente de ingresos fue verificada.
loan_status	El estado actual del préstamo.
pymnt_plan	Indica si se ha establecido un plan de pago para el préstamo.

## 2.

## Metadata

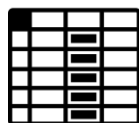
Variable	Descripción
<b>purpose</b>	Una categoría proporcionada por el prestatario para la solicitud del préstamo.
<b>zip_code</b>	Los primeros 3 números del código postal proporcionado por el prestatario en la solicitud del préstamo.
<b>addr_state</b>	El estado proporcionado por el prestatario en la solicitud del préstamo.
<b>delinq_2yrs</b>	El número de incidencias de morosidad de más de 30 días en el archivo de crédito del prestatario en los últimos 2 años.
<b>earliest_cr_line</b>	El mes en que se abrió la primera línea de crédito reportada del prestatario.
<b>inq_last_6mths</b>	El número de consultas en los últimos 6 meses (excluyendo consultas de automóviles e hipotecas).
<b>mths_since_last_delinq</b>	El número de meses desde la última morosidad del prestatario (datos faltantes para algunos registros).
<b>mths_since_last_record</b>	El número de meses desde el último registro público (datos faltantes para algunos registros).
<b>open_acc</b>	El número de líneas de crédito abiertas en el archivo de crédito del prestatario.
<b>pub_rec</b>	Número de registros públicos perjudiciales.
<b>revol_bal</b>	Saldo total de crédito rotativo (es decir, saldo de tarjeta de crédito).
<b>revol_util</b>	Tasa de utilización de la línea rotativa, que representa la cantidad de crédito que el prestatario está utilizando en relación con todo el crédito rotativo disponible.
<b>total_acc</b>	El número total de líneas de crédito actualmente en el archivo de crédito del prestatario.
<b>initial_list_status</b>	El estado de la lista inicial del préstamo, con posibles valores W y F.
<b>application_type</b>	Indica si el préstamo es una solicitud individual o una solicitud conjunta con dos co-prestatarios.
<b>chargeoff_within_12_mths</b>	El número de cancelaciones de cobros en los últimos 12 meses.
<b>pub_rec_bankruptcies</b>	El número de quiebras de registros públicos.
<b>tax_liens</b>	El número de embargos fiscales.

## 2.

# Metadata



**Datos:**  
**39.717**



**Columnas:**  
**111**



**Tasa de interés  
promedio:**  
**11,88%**



**Total de deudores:**  
**5627 morosos**



**Monto prestado  
promedio:**  
U\$D **10681.04**

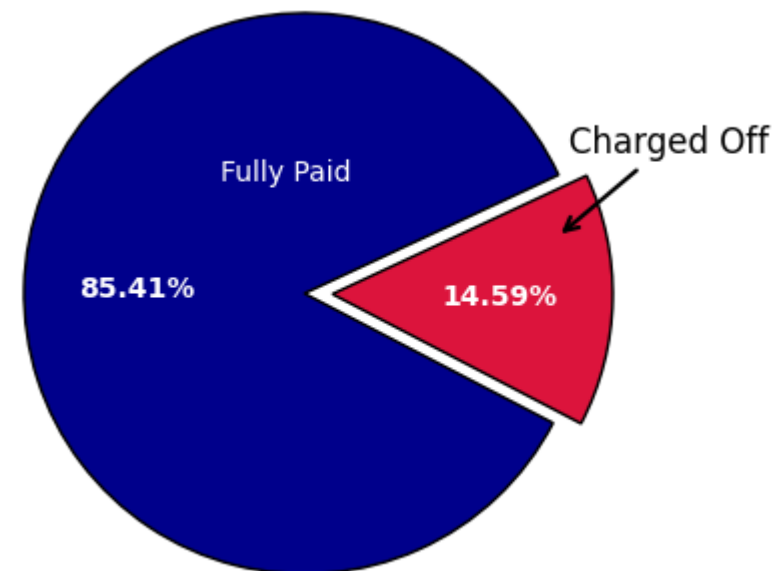


**Monto en riesgo:**  
U\$D **63.968.825 \***

### Resumen:

Este dataset posee gran cantidad de columnas que no nos servirán para el análisis, gran parte se descartará al hacer en análisis exploratorio de datos correspondiente, el proceso se puede apreciar en el archivo .ipynb manejado por Python. Los datos y gráficos empleados se realizaron mediante librerías de este lenguaje de programación.

### Cantidad de deudores



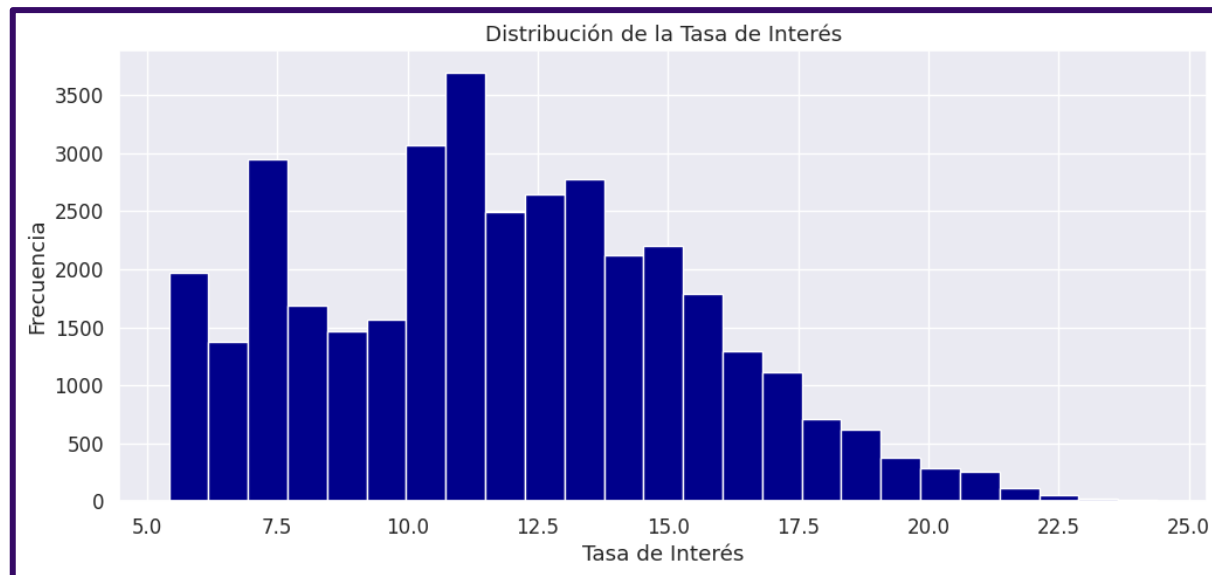
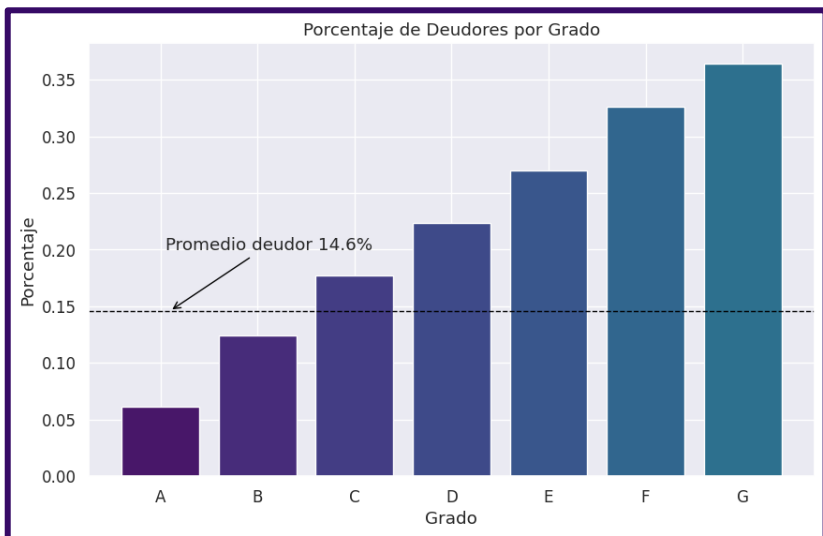
\* Monto hipotético suponiendo que no se pagaran del comienzo



# 3. Análisis Gráfico/Estadístico

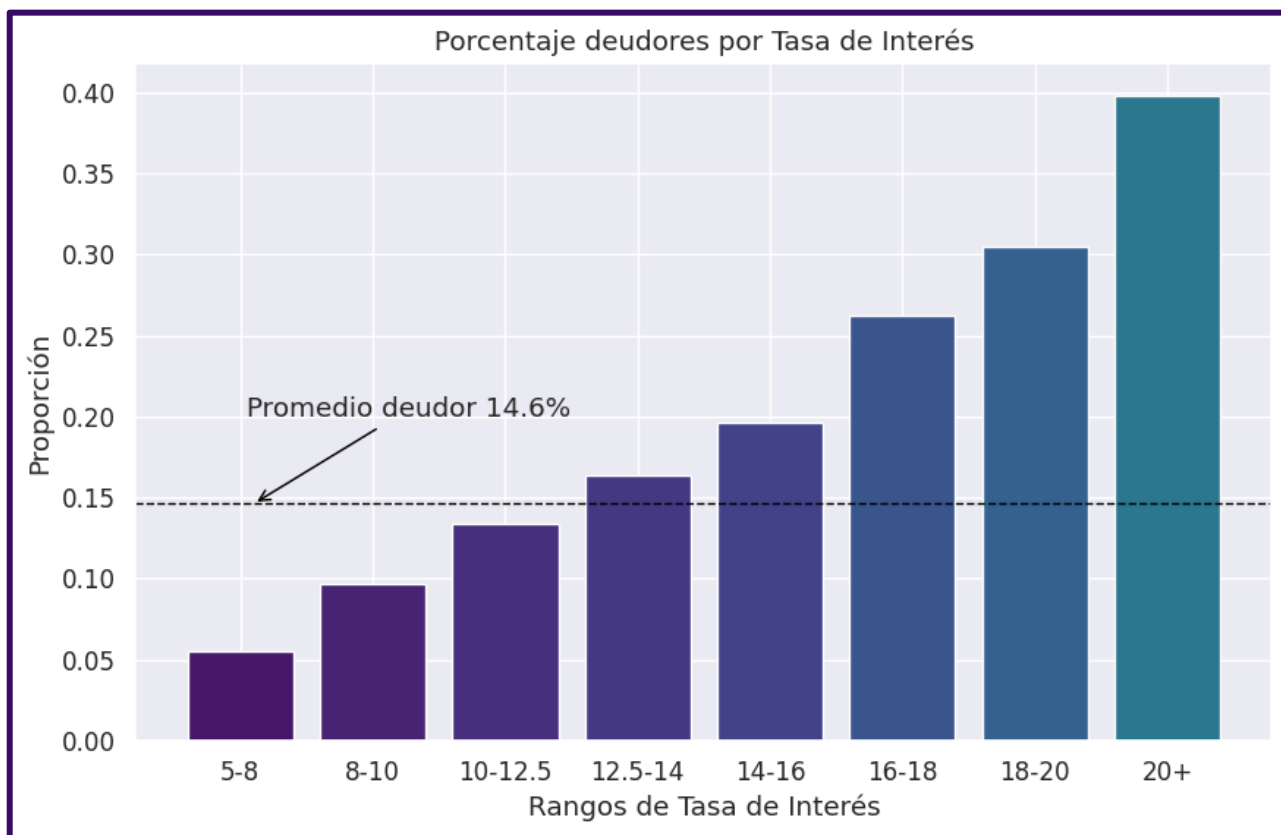
## Análisis de la tasa de interés

Se analizará la composición de la tasa de interés, tanto para los calificados como morosos, como para los que no. También con el grado del prestatario otorgado por el banco.



	Medida estadística	Valor tasa de interes	Interes: morosos	Interes: no morosos
0	Media	11.8879	13.7573	11.5636
1	Moda	10.99	13.49	10.99
2	Varianza	13.4481	13.184	12.783
3	Desvío estándar	3.66716	3.63098	3.57533
4	Asimetría	0.284969	0.0889689	0.312136
5	Curtosis	-0.460522	-0.371335	-0.462932

### 3. Análisis Gráfico/Estadístico



## Efecto de la tasa de interés a la morosidad

En este gráfico se ve el porcentaje de morosidad que hay entre cada rango de tasa de interés. Se puede ver como va aumentando exponencialmente la probabilidad de ser deudor a medida que aumenta la tasa de interés que posee el préstamo. Por ultimo si el préstamo tiene una tasa de 20% o mayor, se puede observar casi un 40% de probabilidades de que haya problema con los pagos, esto quiere decir que, de cada 5 prestamos que se da con esta tasa, 2 van a convertirse en morosos y no terminar de pagar el préstamo.

# 4. Modelización

Luego del manejo de los datos que incluyo:

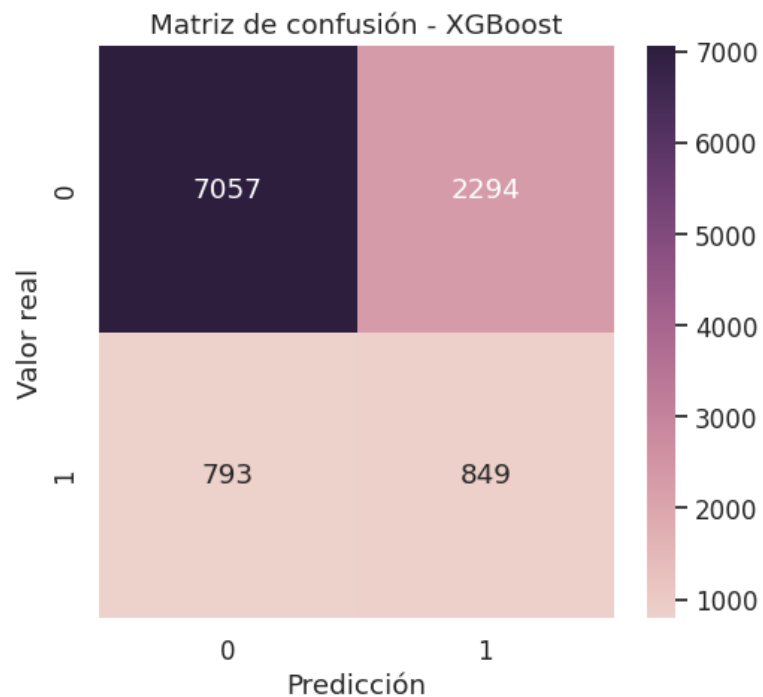
- Eliminación de valores nulos y outliers
- Eliminación de columnas redundantes.
- Transformación de los datos categóricos en valores numéricos.
- Aplicación de técnicas como OneHotEncoder y LabelEncoder.
- Aplicación de técnicas de OverSampling

Se aplicó los siguientes **modelos predictivos**:

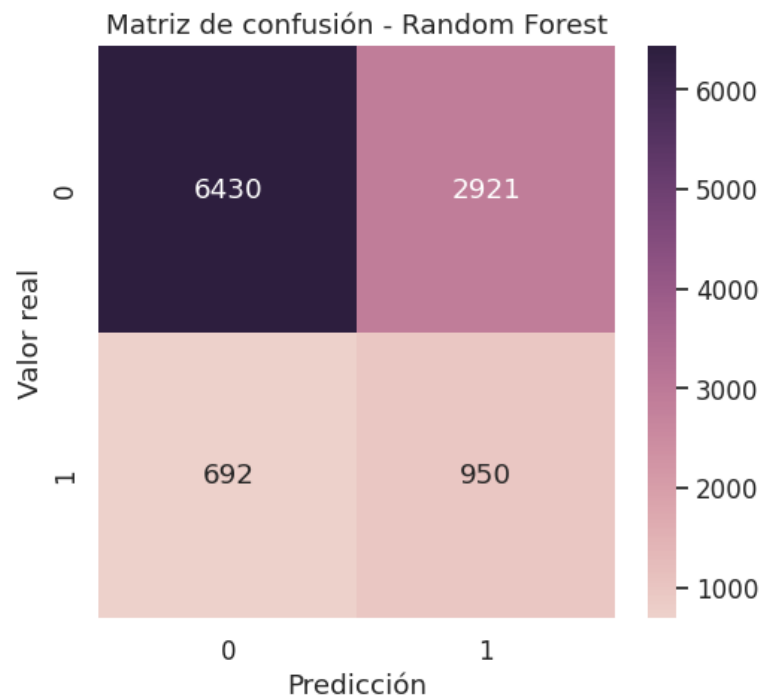
- **XGBoost (Extreme Gradient Boosting)**: Potente técnica de aprendizaje automático basada en Gradient Boosting. Es altamente eficiente y efectiva en problemas de clasificación y regresión, especialmente con grandes conjuntos de datos y clases desequilibradas.
- **Random Forest**: Algoritmo de ensamblaje de árboles de decisión que mejora la precisión al combinar múltiples árboles. Es versátil, fácil de usar y se destaca en manejar datos numéricos y categóricos.
- **Regresión Logística**: Modelo de clasificación lineal utilizado para problemas binarios o categóricos. Calcula la probabilidad de pertenecer a una clase y es eficiente y simple en su implementación.

## 4.

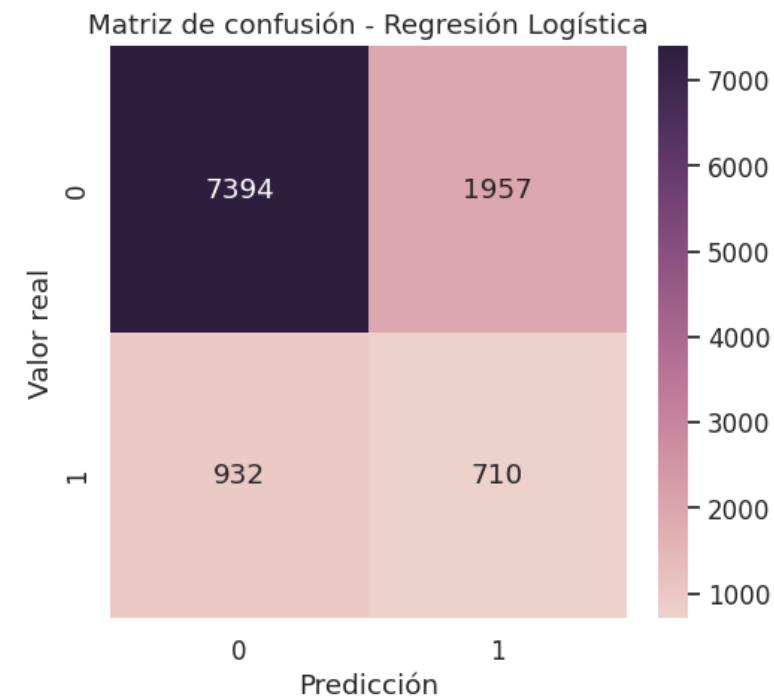
## Modelización: Resultados



XGBoost	
Accuracy	0,7191
F1 Score	0,7509



Random Forest	
Accuracy	0,6713
F1 Score	0,7155



Regresión Logística	
Accuracy	0,732
F1 Score	0,761

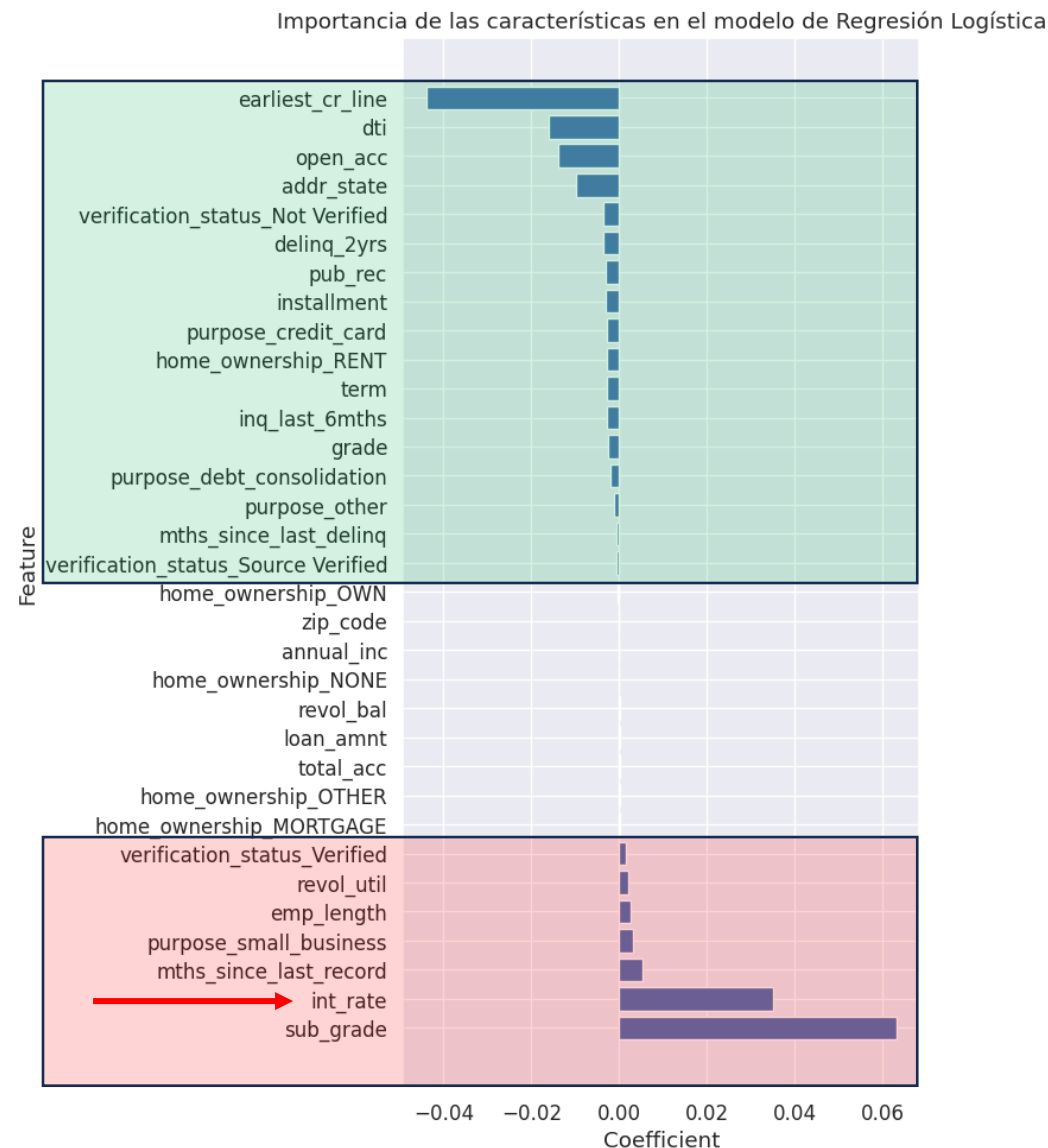
# 4. Modelización: Resultados

Por los resultados anteriores, principalmente haciendo foco en el F1 Score, ya que el accuracy al ser datos desbalanceados no es un buen indicador, se usará la **Regresión Logística** como modelo para predecir morosidad.

En el gráfico de nuestra derecha se aprecia cuáles son las variables que más afectan el resultado:

- **En verde:** Variables donde al tener valor más alto, hay mayor probabilidad de que no tenga problemas con los pagos.
- **En rojo:** Variables que, al poseer un valor más alto, hay más posibilidad que el préstamo se convierta en deudor.

Entre estas últimas se encuentra la **tasa de interés** y el **Sub-Grado** del préstamo, por lo que se puede apreciar lo tanto que afecta la tasa de interés para que haya problema al devolver el dinero, ya que Sub-Grade también es afectado por la tasa, a mayor grado, mayor interés tendrá que pagar.





## 4. Conclusión

Por ultimo, haciendo uso del modelo, podemos crear un hipotético caso donde los datos tengan una mayor tasa de interés para ver como afectará al porcentaje de morosidad:

	Porcentaje deudor	Aumento/Disminución
DataFrame original	14.59%	-
Regresión Logistica	24.26%	9.67%
RL con +10% de interés	25.96%	+1.7%

Aplicamos un aumento a los datos original de un 10% a todas las tasa de interés, por lo que este modelo termina arrojando un aumento del 1.7% de la morosidad, lo que equivale a unos 676 prestamos, multiplicados por el monto de préstamo promedio nos da una perdida total de **\$7.211.693,71**. Por lo que se aprecia el efecto que tiene un posible aumento general de las tasas de interés global.