

Explicación descriptiva:

Los bancos obtienen importantes ingresos de la concesión de préstamos. Pero a menudo va asociado a un riesgo. El prestatario puede no pagar el préstamo. Para mitigar este problema, los bancos han decidido utilizar el aprendizaje automático. Han recopilado datos anteriores sobre los prestatarios de préstamos y les gustaría desarrollar un sólido modelo de ML para clasificar si es probable que un nuevo prestatario incumpla o no.

El conjunto de datos es enorme y consta de múltiples factores deterministas como los ingresos del prestatario, el sexo, la finalidad del préstamo, etc. El conjunto de datos está sujeto a una fuerte multicolinealidad y a valores vacíos.

Preguntas/Hipótesis:

¿Existe una relación significativa entre el nivel de ingresos del prestatario y la probabilidad de incumplimiento del préstamo?

Hipótesis: Los prestatarios con ingresos más bajos tienen una mayor probabilidad de incumplir con el pago del préstamo en comparación con aquellos con ingresos más altos.

¿La finalidad del préstamo influye en la probabilidad de incumplimiento?

Hipótesis: Los préstamos destinados a gastos discrecionales (por ejemplo, viajes o compras de lujo) tienen una mayor tasa de incumplimiento que los préstamos destinados a necesidades básicas o inversiones (por ejemplo, educación o vivienda).

¿La multicolinealidad entre las variables económicas del prestatario afecta la precisión del modelo predictivo?

Hipótesis: La multicolinealidad entre factores como ingresos, deudas y otros indicadores financieros reduce la capacidad del modelo para predecir con precisión el incumplimiento.

¿Los datos demográficos (sexo, edad, estado civil) tienen un impacto significativo en la probabilidad de incumplimiento del préstamo?

Hipótesis: Algunas características demográficas, como la edad o el estado civil, pueden ser predictores significativos de la probabilidad de incumplimiento.