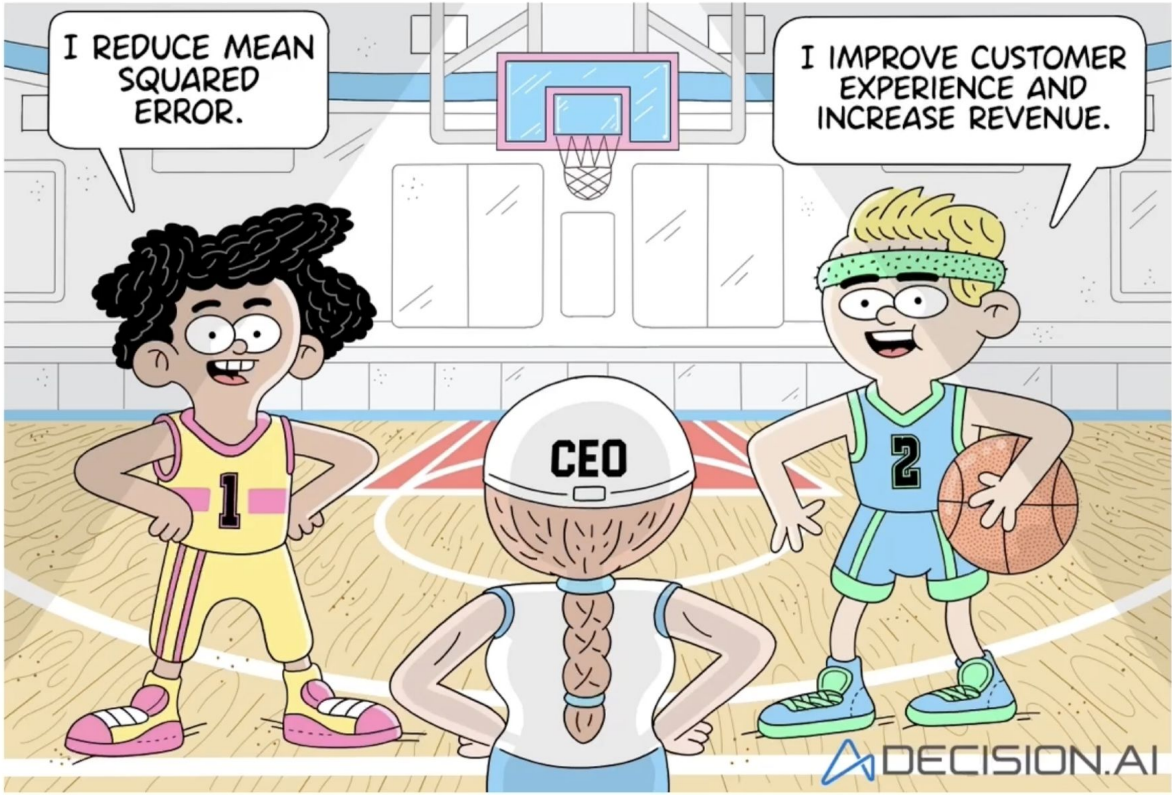
Conectando métricas de negocio con nuestros modelos de ML.

Por: Jose Alberto Arango S.

¿Cuál DS crees que ella elegirá?



Problema

Standard Churn Model Scenario

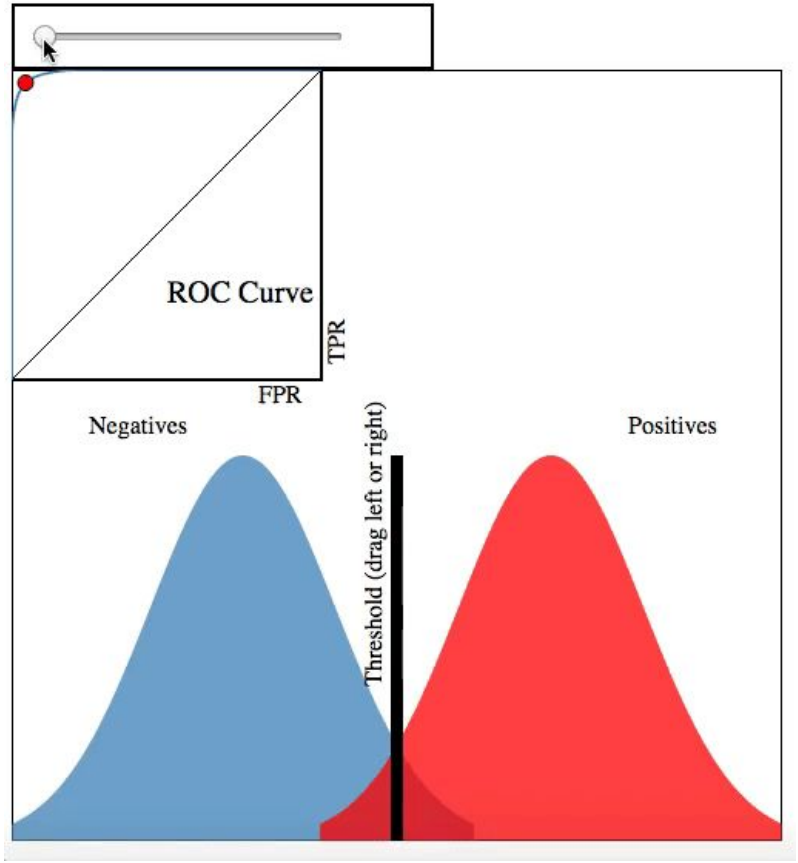
Churn	Contract	Years as Customer	Home Phone
0	Monthly	7	0
0	Annual	1	0
1	Monthly	3	1
0	Monthly	3	0

Prediction Target

Course Outline

Prediction	Churn: Yes	Churn: No
Yes	150,000	80,000
No	20,000	3,000,000

¿Cómo escogemos el umbral de decisión?



¿Cómo escogemos el umbral de decisión?

```
threshold = 0.5
y_pred_binary  = np.where(y_pred >= threshold, 1, 0)
```

Profit Curve!

Medir el rendimiento de nuestros modelos ML en términos de financieros ($).

Profit matrix

Profit matrix

		Sale	
		Yes	No
Call	True	$5	-$2
	False	$0	$0

Profit matrix

	Collection		
Call		**Yes**	**No**
	True	$10	-$2
	False	$0	$0

Traducir los resultados en impacto financiero

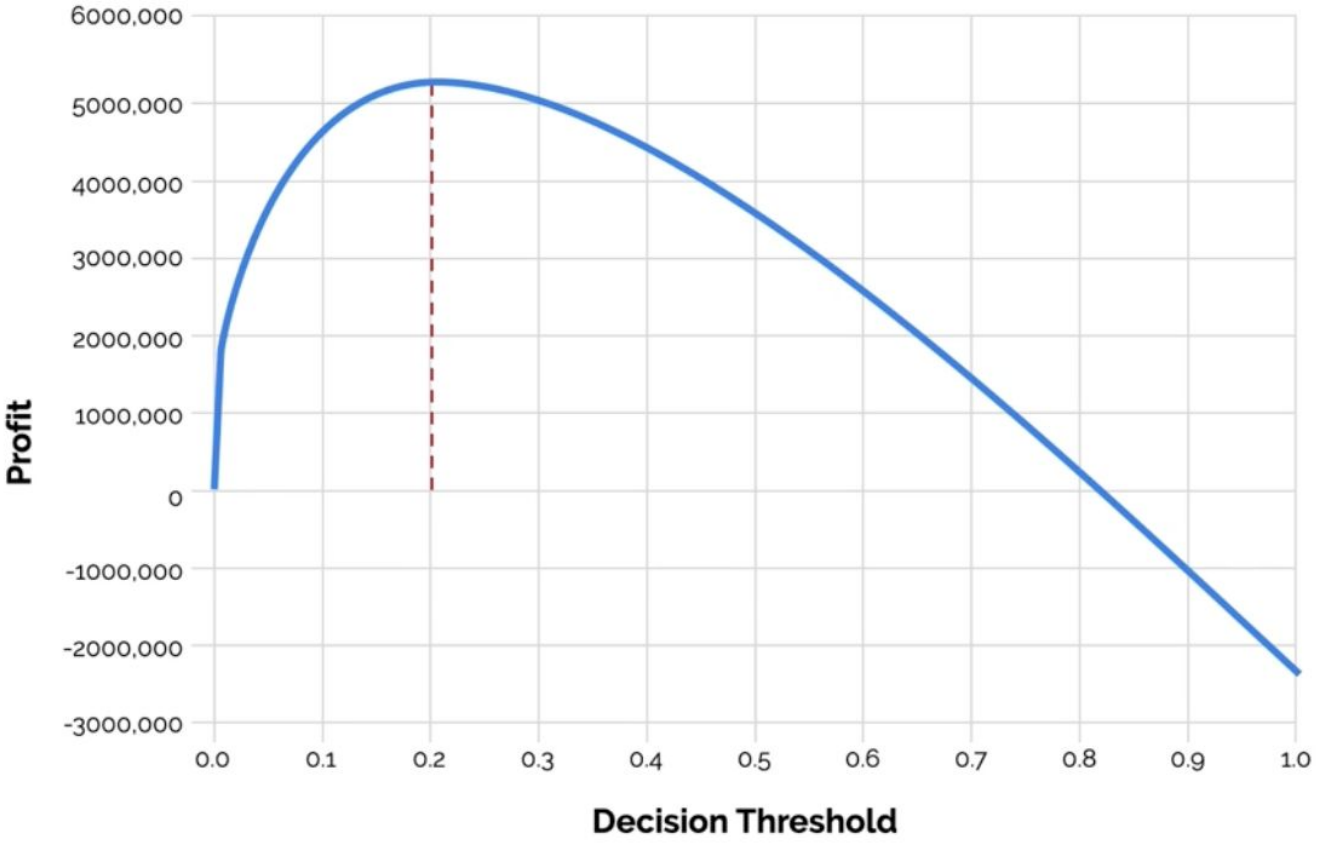
Discount	Churned: Yes	Churned: No
True	150,000	80,000
False	20,000	3,000,000

Discount	Churned: Yes	Churned: No
True	$200	-$30
False	$0	$0

150,000 * 200
- 80,000 * 30

= **$27,600,000**

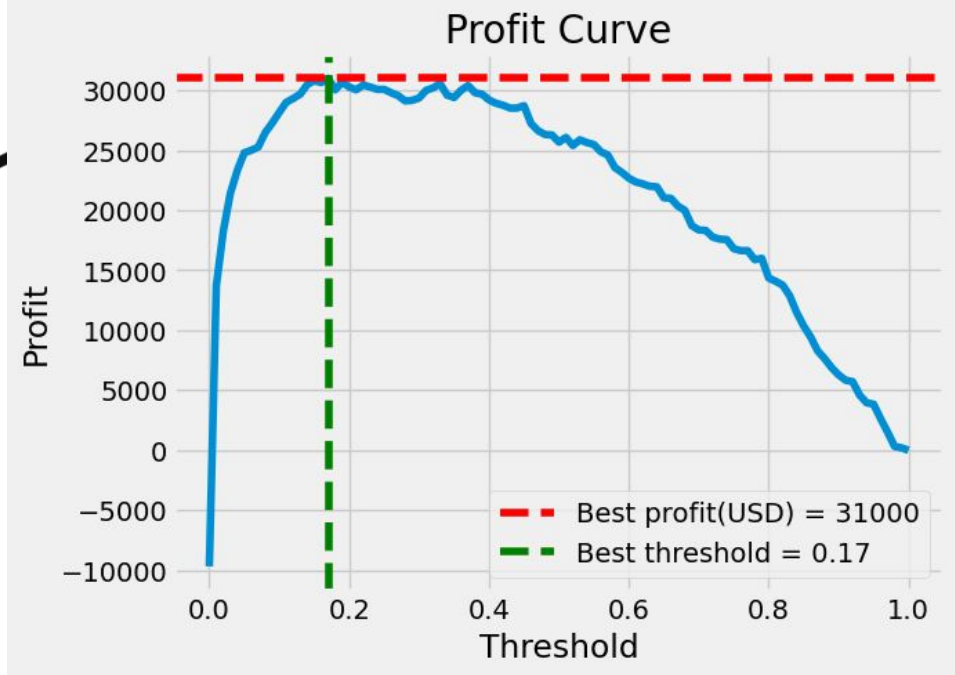
Profit curve

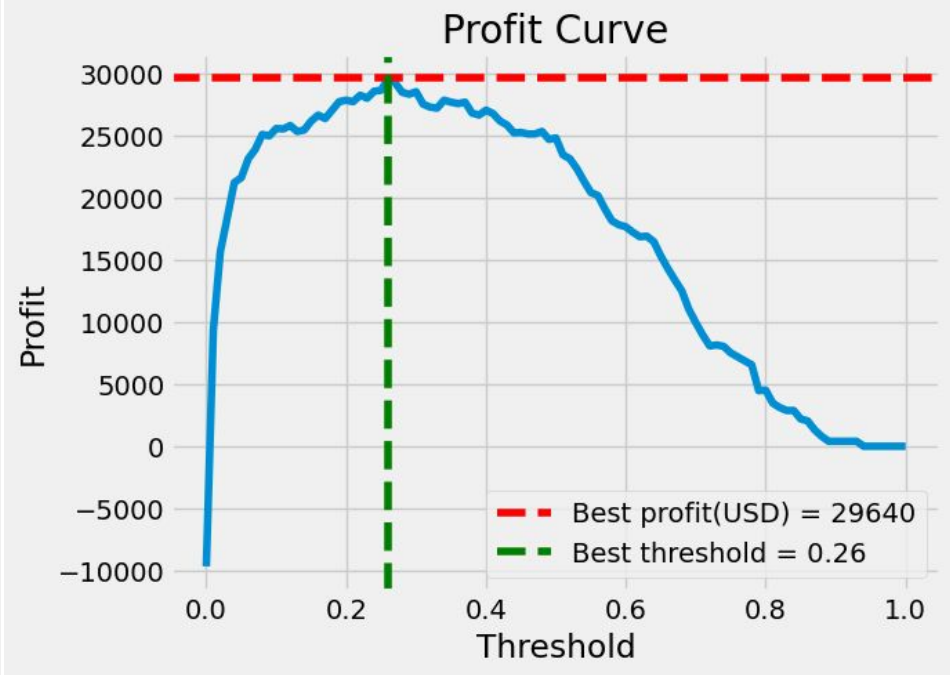


Limitaciones:

- No saber los valores de los TP y FP
- Solo problemas de clasificación
- Valores heterogéneos

La importancia de una variables





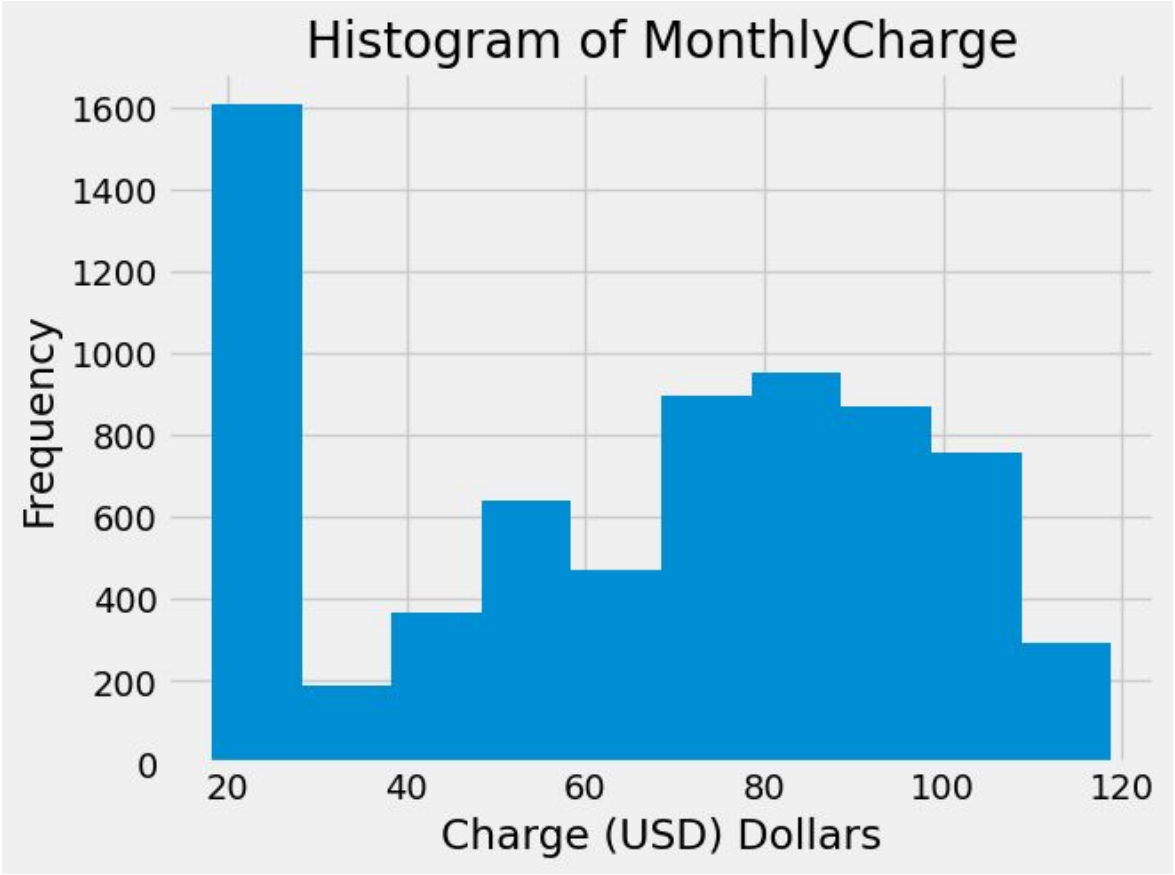
Todas las variables

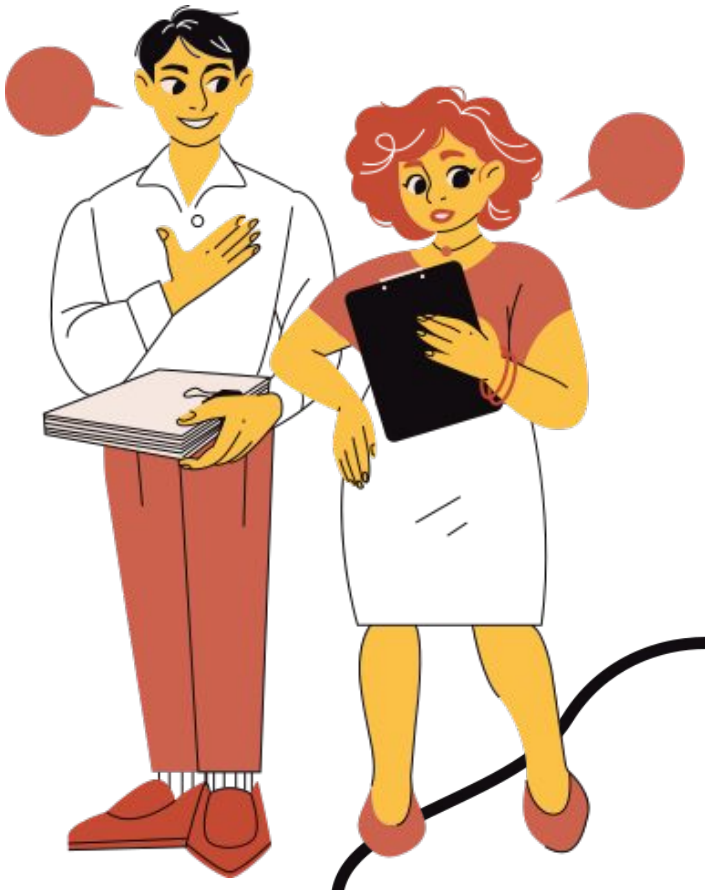
Falta 1 variable

Total pérdida: 1.360 USD

Más allá de los umbrales de decisión

Diferentes grupos de clientes

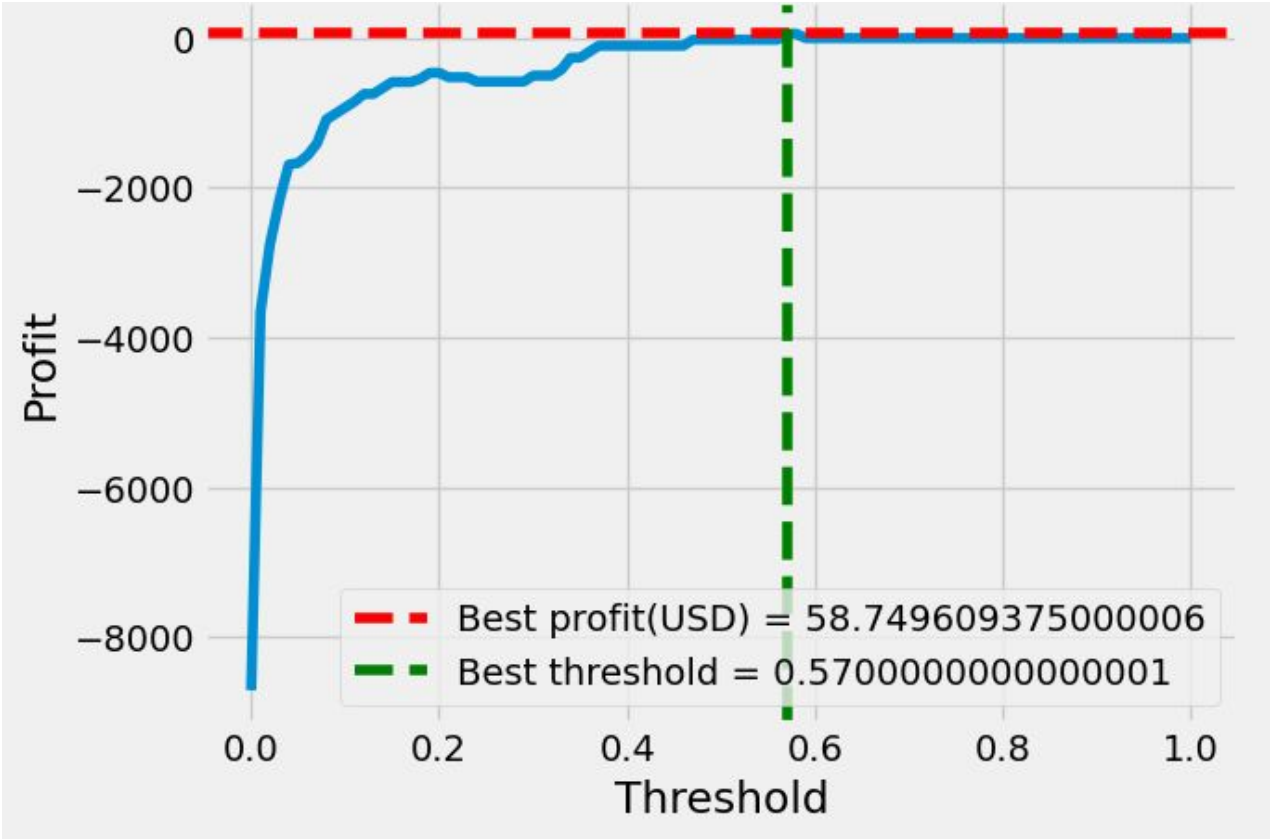




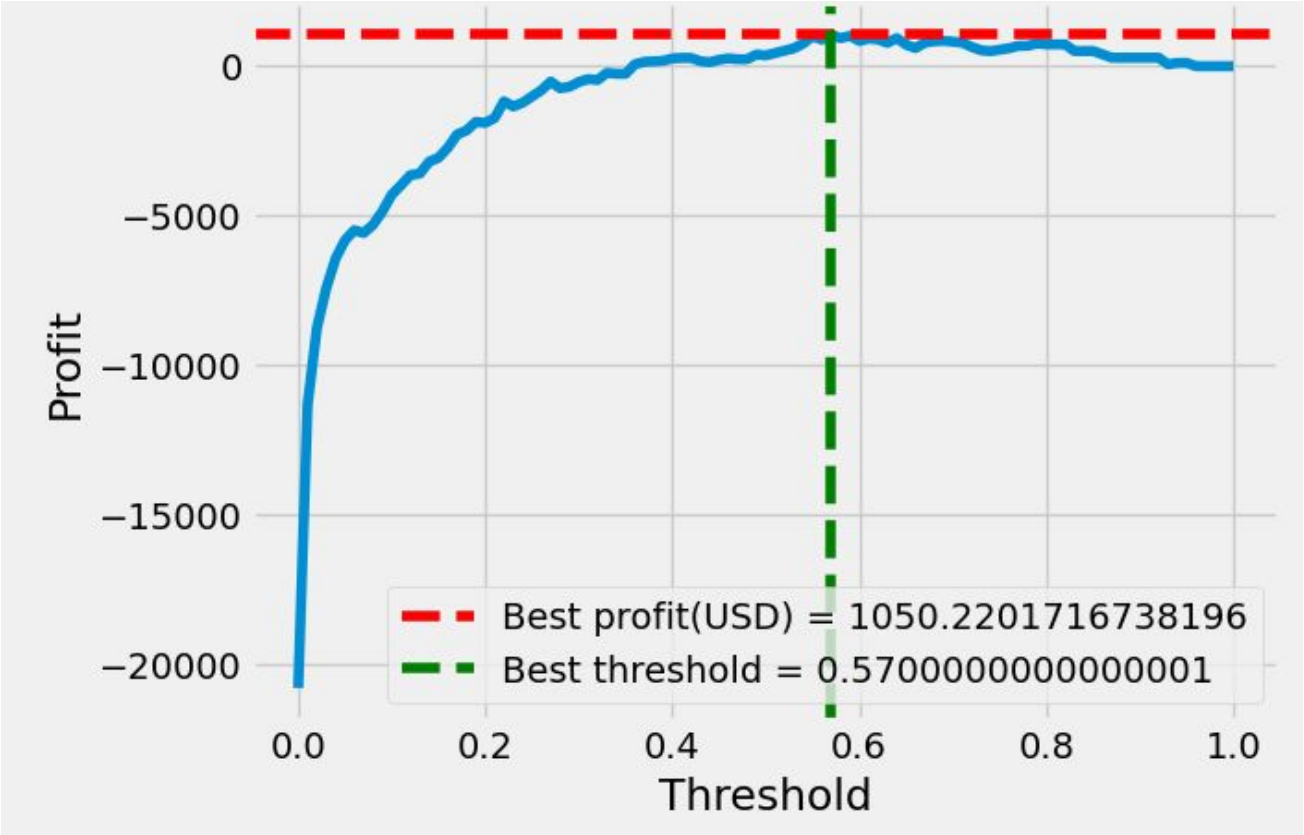
Profit matrix

Discount	Churned: Yes	Churned: No
True	$3*mean_monthly_charge	-$80
False	$0	$0

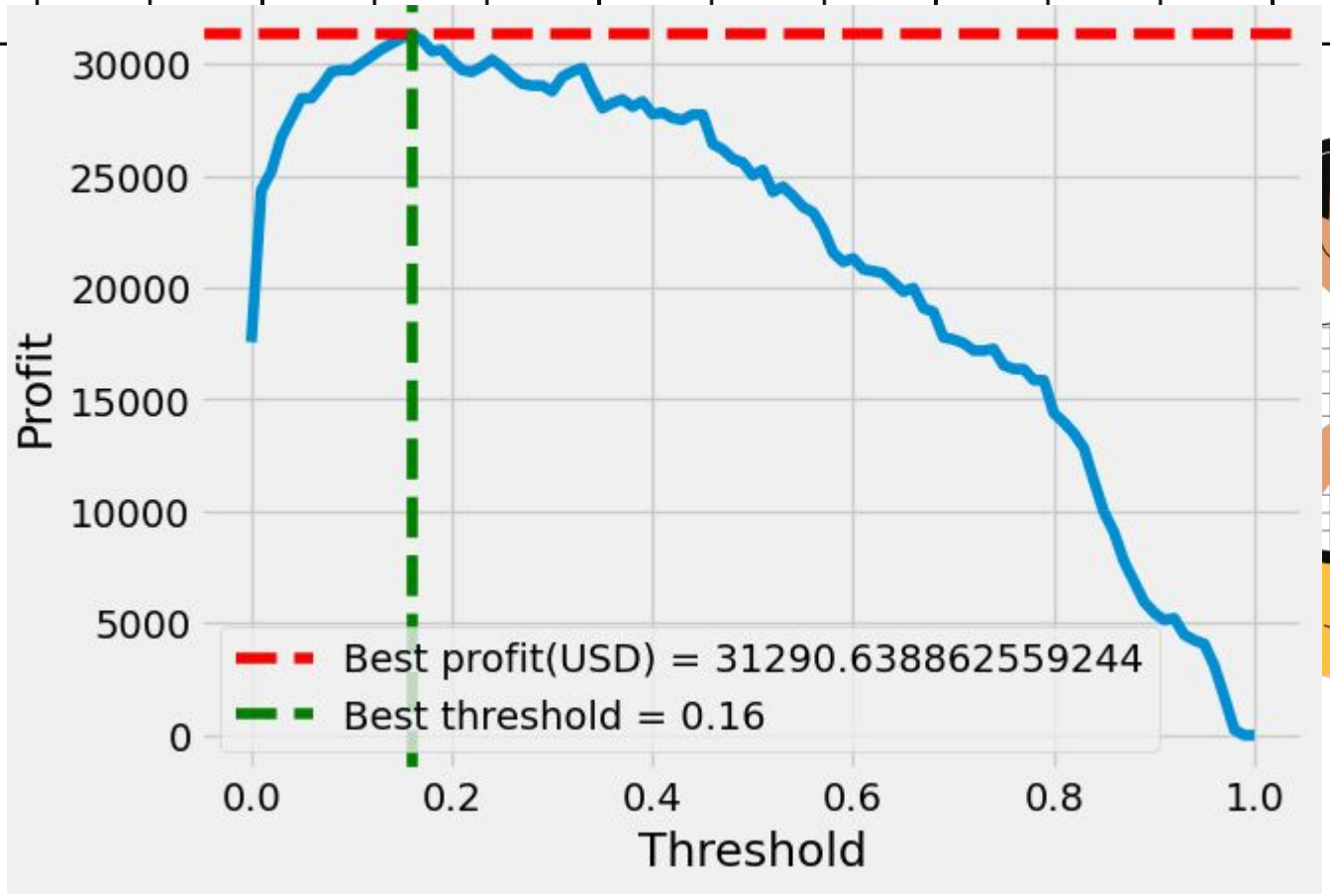
Umbral para clientes que pagan menos de $20 al mes



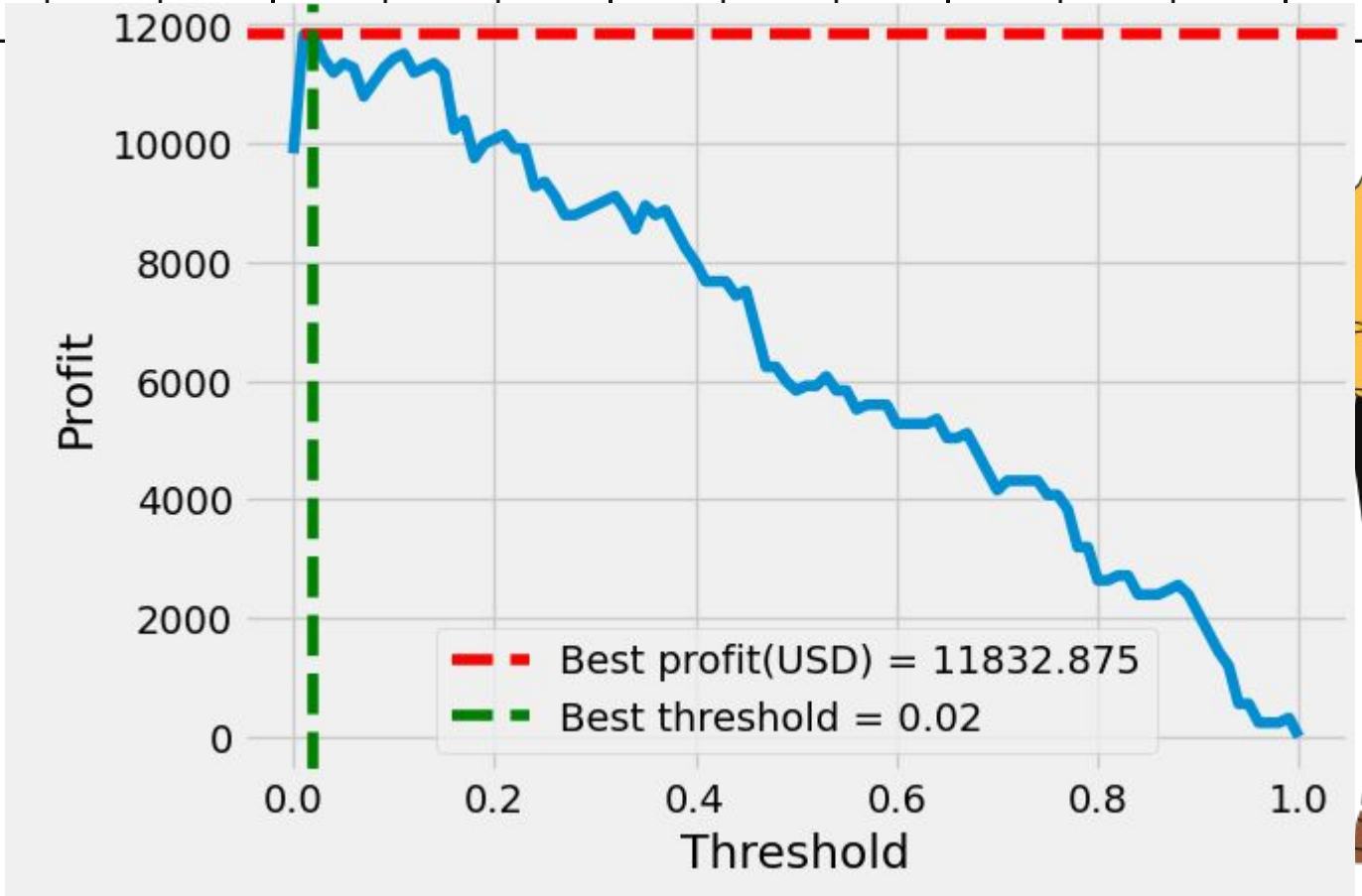
Umbral para clientes que pagan entre $20 y $60/mes



Umbral para clientes que pagan entre $60 y $100/mes

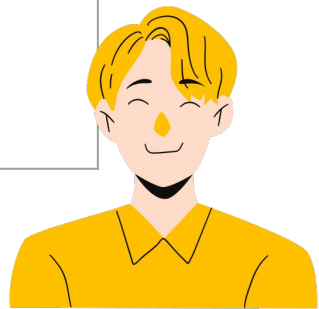


Umbral para clientes que pagan más de $100/mes



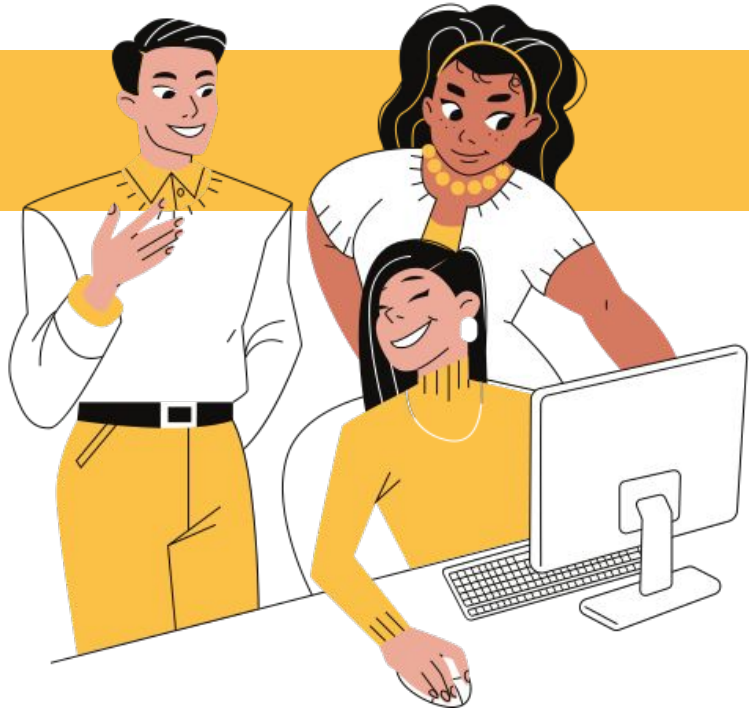
Tipos de umbral

Tipo de Umbral	Ganancia esperada
Umbrales flexibles	**44.232 USD**
Umbral estático	**31.000 USD**
Diferencia	**13.232 USD**



Mejoras del modelo

¿Debo seguir mejorando este modelo o pasar a un nuevo proyecto?



“Datos esperados” que sean precisos.

```
errors = y_pred - y_true
```

“Datos esperados” que sean precisos.

```
synth_data = y_pred - errors * improvement_factor
```

“Datos esperados” que sean precisos.

```
synth_data_AUC = roc_auc_score(y_true, synth_data)
```

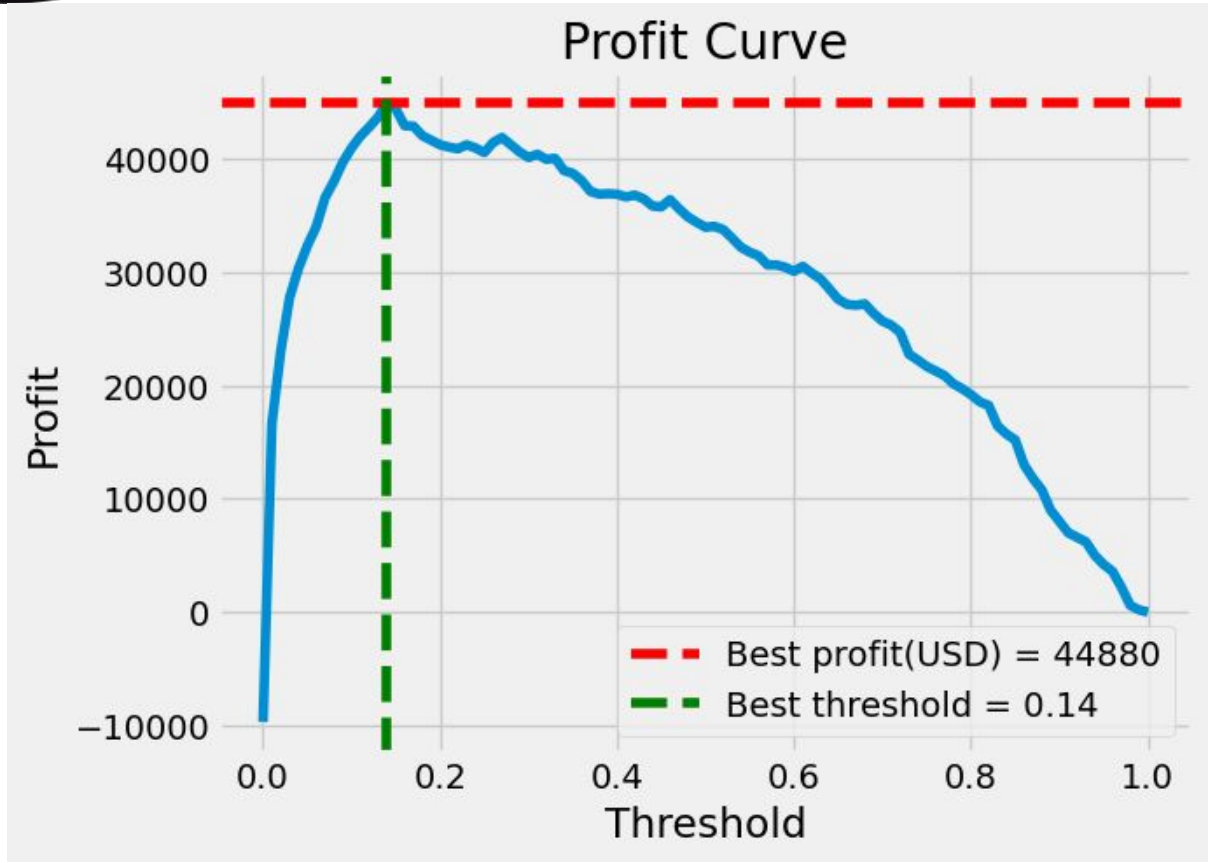
```python
def make_synth_accurate_data(y_pred, y_true, improvement_factor=0.1):
    """Create synthetic data that is accurate.

    Args:
        y_pred (array-like): Predicted probabilities.
        y_true (array-like): True labels.
        improvement_factor (float): A measure of how much to increase the accuracy of the synthetic data.

    Returns:
        array: Synthetic data.
    """
    errors = y_pred - y_true
    synth_data = y_pred - errors * improvement_factor
    synth_data_AUC = roc_auc_score(y_true, synth_data)
    print(
        f"improvement_factor: {improvement_factor:.2f}.   AUC of synthetic data: {synth_data_AUC:.2f}"
    )
    return synth_data
```

Diferentes valores de Improvement_factor

```
for i in np.linspace(0, 0.2, 11):
    make_synth_accurate_data(y_pred, y_test, improvement_factor=i)
```

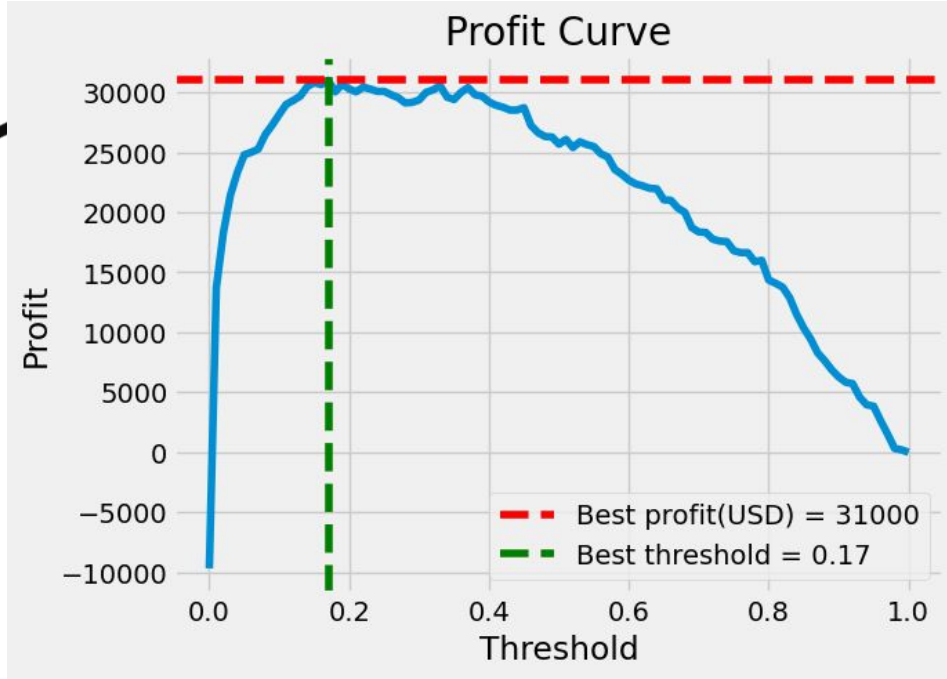
```
improvement_factor: 0.00.   AUC of synthetic data: 0.81
improvement_factor: 0.02.   AUC of synthetic data: 0.83
improvement_factor: 0.04.   AUC of synthetic data: 0.84
improvement_factor: 0.06.   AUC of synthetic data: 0.85
improvement_factor: 0.08.   AUC of synthetic data: 0.86
improvement_factor: 0.10.   AUC of synthetic data: 0.87
improvement_factor: 0.12.   AUC of synthetic data: 0.88
improvement_factor: 0.14.   AUC of synthetic data: 0.89
improvement_factor: 0.16.   AUC of synthetic data: 0.90
improvement_factor: 0.18.   AUC of synthetic data: 0.91
improvement_factor: 0.20.   AUC of synthetic data: 0.92
```

Profit curve esperada con la mejora

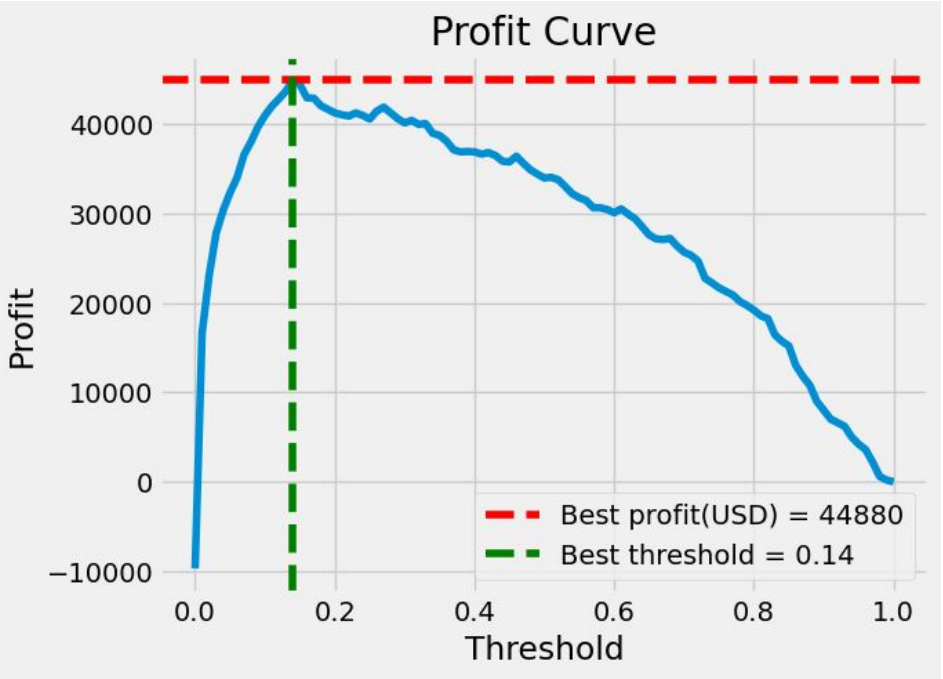


- **Improvement factor: 0.14.**
- **AUC of synthetic data: 0.89**

Profit curve Actual vs Esperado



Actual



Esperado

Total ganancia: 13.880 USD

Referencias

1. Dataset usado: Telco Customer Churn
2. Machine Learning for Business Decision Optimization W&B AI Academy
3. Decision.AI

Muchas Gracias!!!

¿Alguna pregunta?