

Optimización Estratégica en Instituciones Financieras

Identificación y Predicción de Fuga de Clientes
mediante Aprendizaje Automático

Eduardo David González

Subdirector de Analítica Avanzada & ML-Ops, Grupo Regional

Datos y metodología

customer_id	credit_score	country	gender	age	tenure	balance	products_number	credit_card	active_member	estimated_salary	churn
15634602	619	France	Female	42	2	0	1	1	1	101348.88	1
15647311	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
15619304	502	France	Female	42	8	159660.8	3	1	0	113931.57	1
15701354	699	France	Female	39	1	0	2	0	0	93826.63	0
15737888	850	Spain	Female	43	2	125510.82	1	1	1	79084.1	0
15574012	645	Spain	Male	44	8	113755.78	2	1	0	149756.71	1
15592531	822	France	Male	50	7	0	2	1	1	10062.8	0
15656148	376	Germany	Female	29	4	115046.74	4	1	0	119346.88	1
15792365	501	France	Male	44	4	142051.07	2	0	1	74940.5	0
15592389	684	France	Male	27	2	134603.88	1	1	1	71725.73	0
15767821	528	France	Male	31	6	102016.72	2	0	0	80181.12	0
15737173	497	Spain	Male	24	3	0	2	1	0	76390.01	0
15632264	476	France	Female	34	10	0	2	1	0	26260.98	0
15691483	549	France	Female	25	5	0	2	0	0	190857.79	0
15600882	635	Spain	Female	35	7	0	2	1	1	65951.65	0

Análisis descriptivo

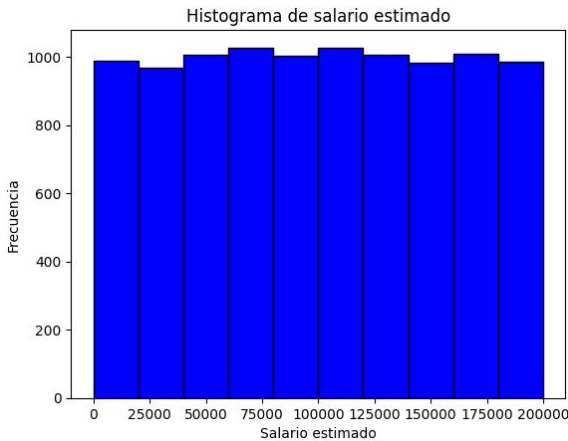
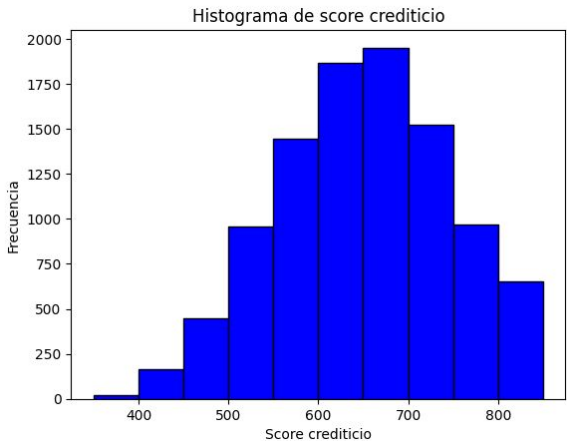
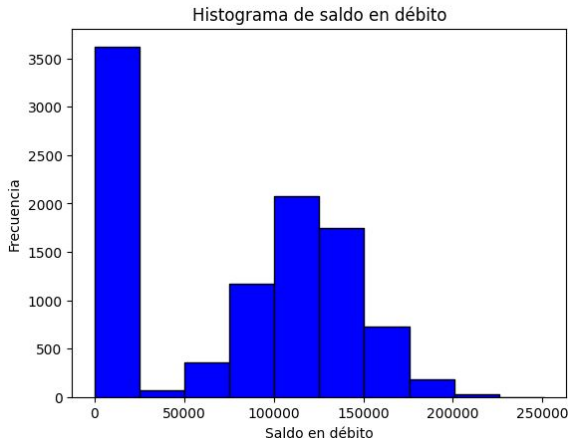
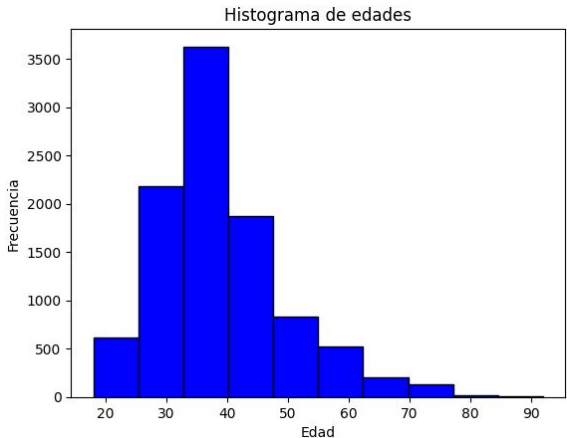


Tabla 1 Distribución de género

Género	Registros
Male	5457
Female	4543

Tabla 2 Distribución de países

País	Registros
Francia	5014
Alemania	2509
España	2477

Tabla 3 Distribución de antigüedad de los clientes (años)

Antigüedad	Registros
0	413
1	1035
3	1009
4	989
5	1012
6	967
7	1028
8	1025
9	984
10	490

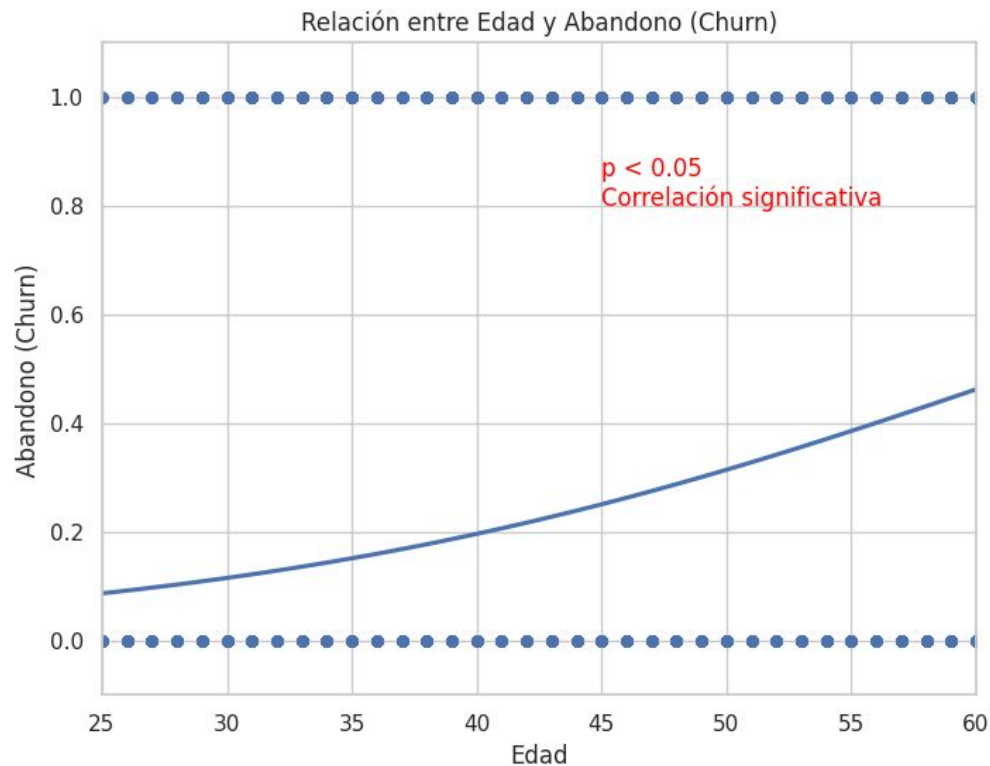
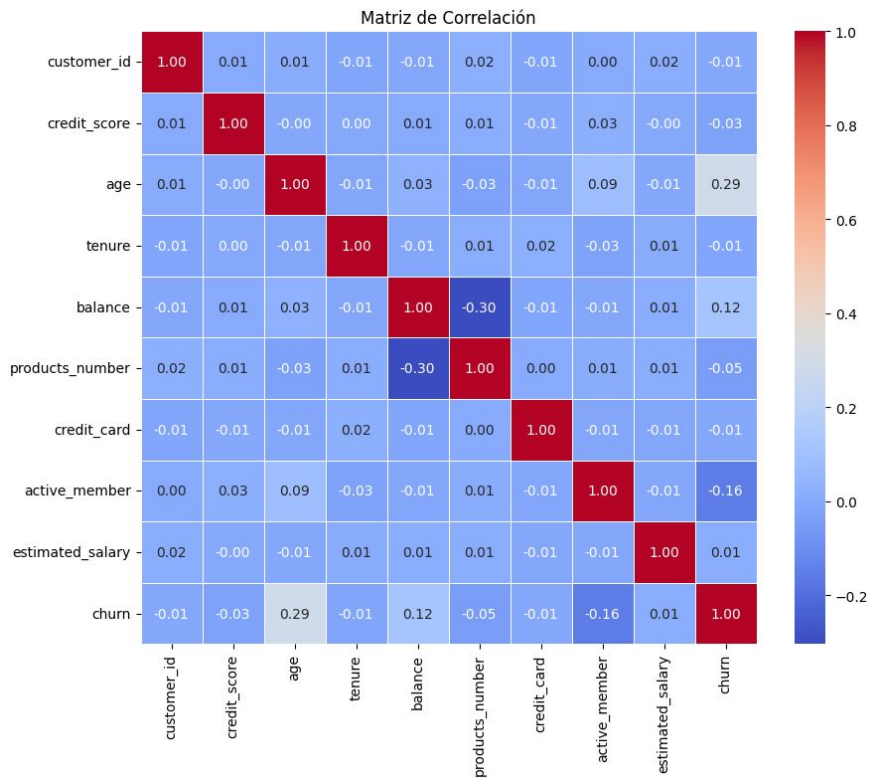
Tabla 4 Distribución de cantidad de productos contratados

Productos	Registros
1	5084
2	4590
3	266
4	60

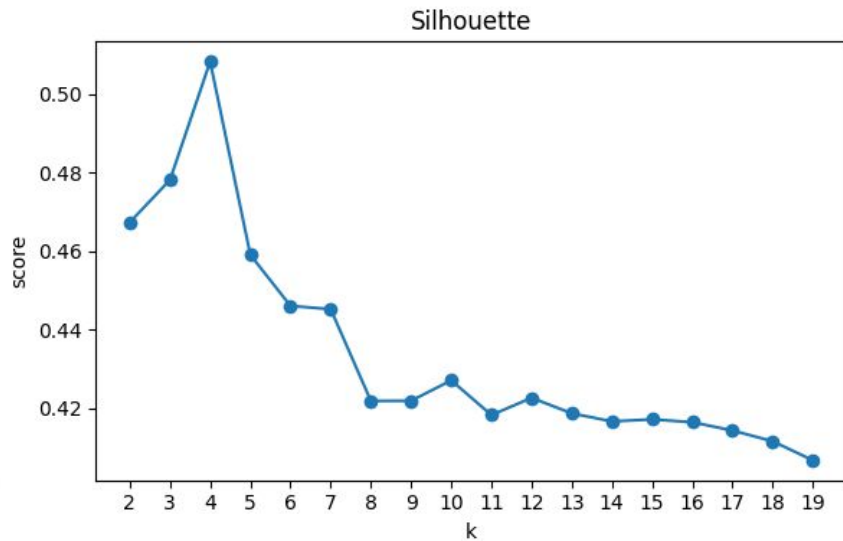
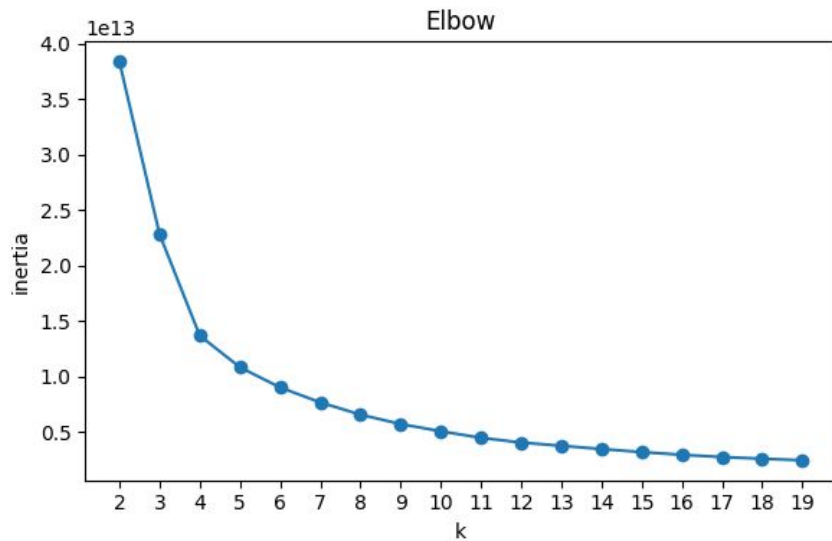
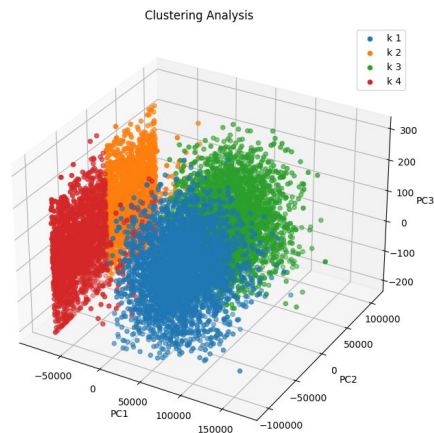
Tabla 5 Distribucion de tenencia de tarjeta de crédito (tdc)

tdc	Registros
Si	7055
No	2945

Correlación



Análisis de clustering

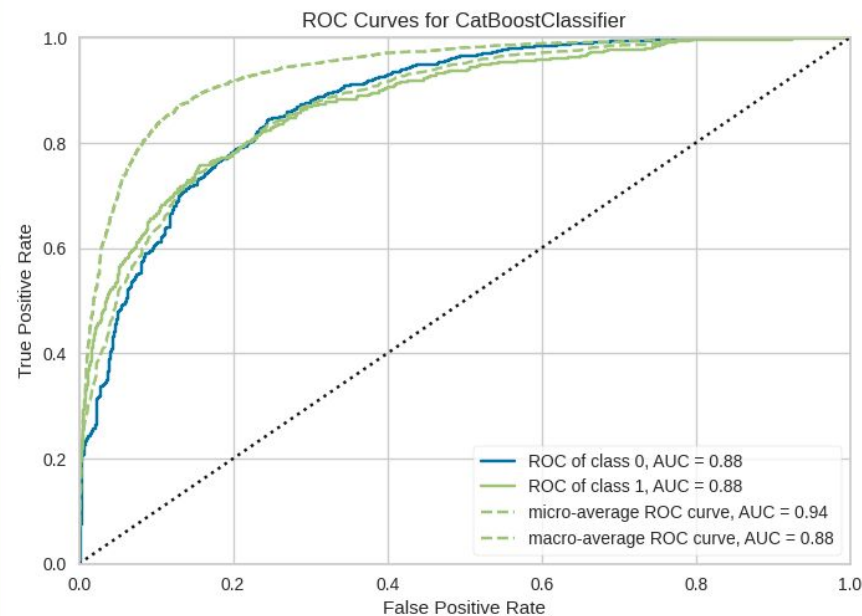
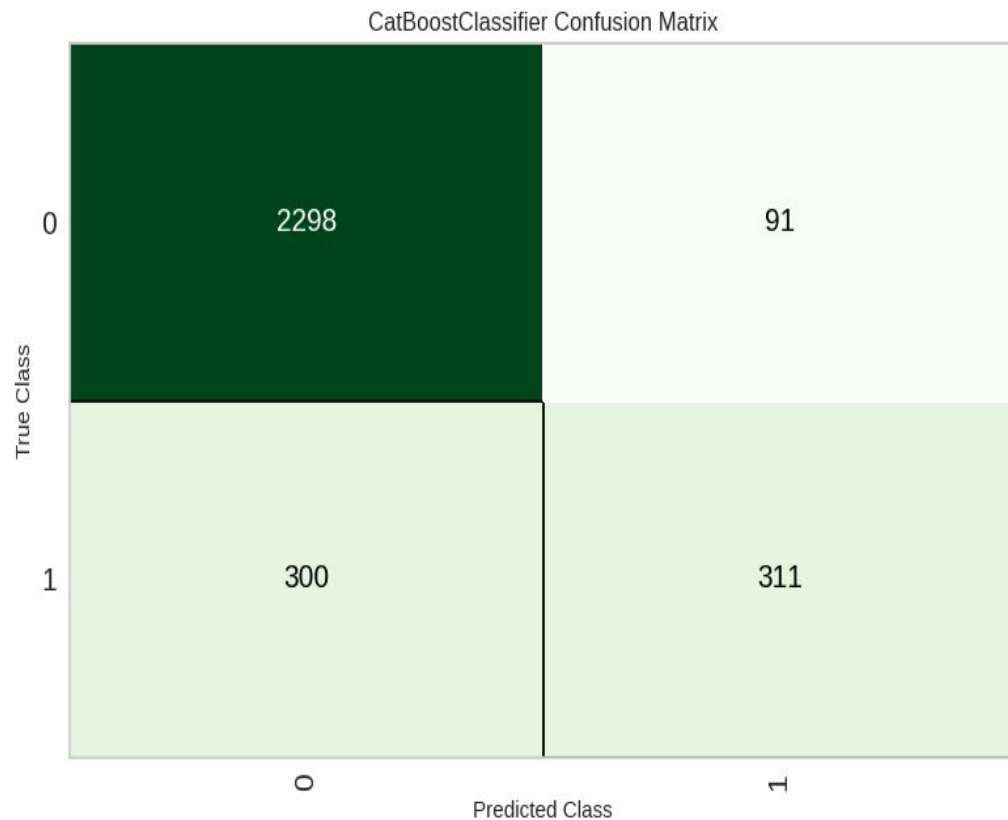


Modelado predictivo

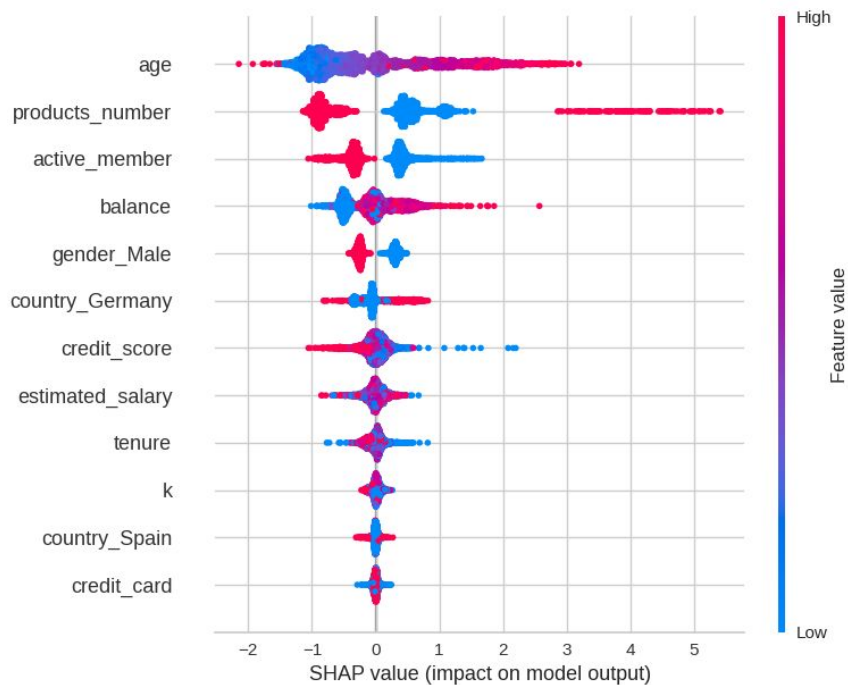
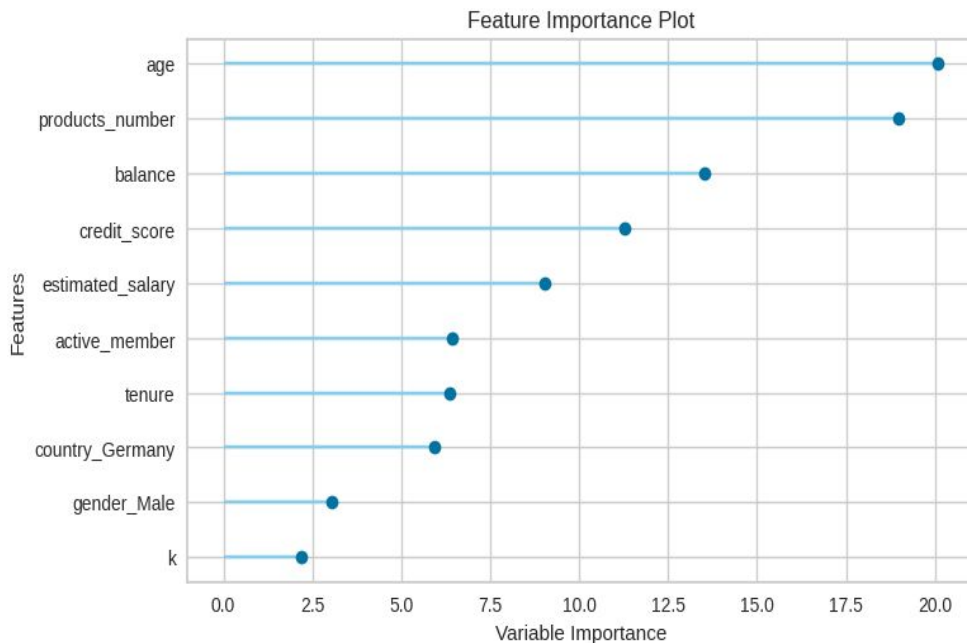
Tabla 7 Comparación de Rendimiento de Modelos

Modelo	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC	TT (Sec)
catboost	0.8616	0.8571	0.4818	0.7516	0.5860	0.5077	0.5265	3.7690
gbc	0.8621	0.8618	0.4650	0.7693	0.5787	0.5023	0.5255	0.1220
rf	0.8584	0.8430	0.4432	0.7638	0.5605	0.4831	0.5088	0.1270
ada	0.8531	0.8429	0.4552	0.7238	0.5578	0.4753	0.4942	0.0490
xgboost	0.8471	0.8283	0.4812	0.6759	0.5614	0.4721	0.4825	0.0730
et	0.8501	0.8380	0.4208	0.7298	0.5333	0.4517	0.4763	0.0980
qda	0.8246	0.8055	0.3878	0.6118	0.4743	0.3752	0.3895	0.0060
dt	0.7863	0.6746	0.4860	0.4761	0.4806	0.3462	0.3465	0.0100
lda	0.8106	0.7726	0.2504	0.5837	0.3490	0.2587	0.2911	0.0090
ridge	0.8079	0.0000	0.1346	0.6354	0.2215	0.1620	0.2269	0.0080
nb	0.7873	0.7486	0.1003	0.4154	0.1606	0.0880	0.1183	0.0070
lr	0.7911	0.6813	0.0743	0.4283	0.1261	0.0704	0.1063	0.5230
knn	0.7606	0.5286	0.0806	0.2403	0.1204	0.0201	0.0243	0.1930
dummy	0.7963	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0070
svm	0.7487	0.0000	0.0540	0.0562	0.0412	-0.0152	-0.0158	0.0160

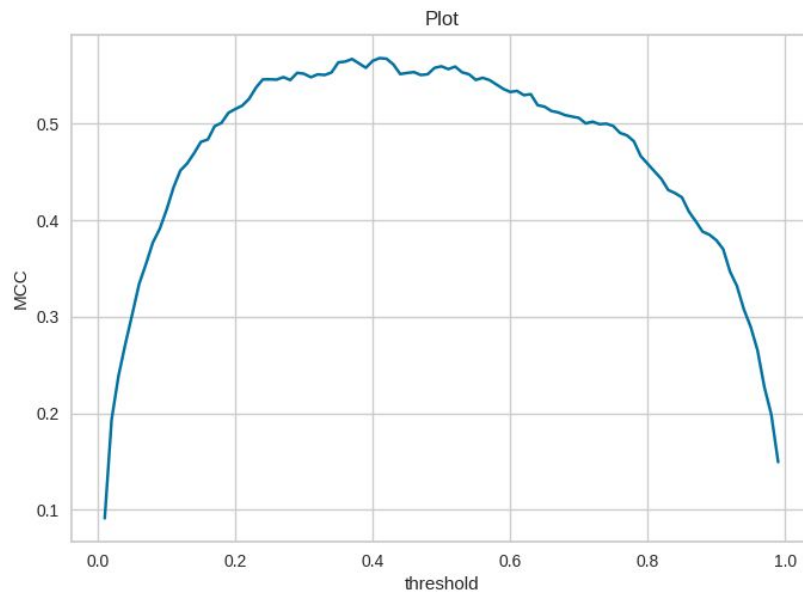
Modelado predictivo



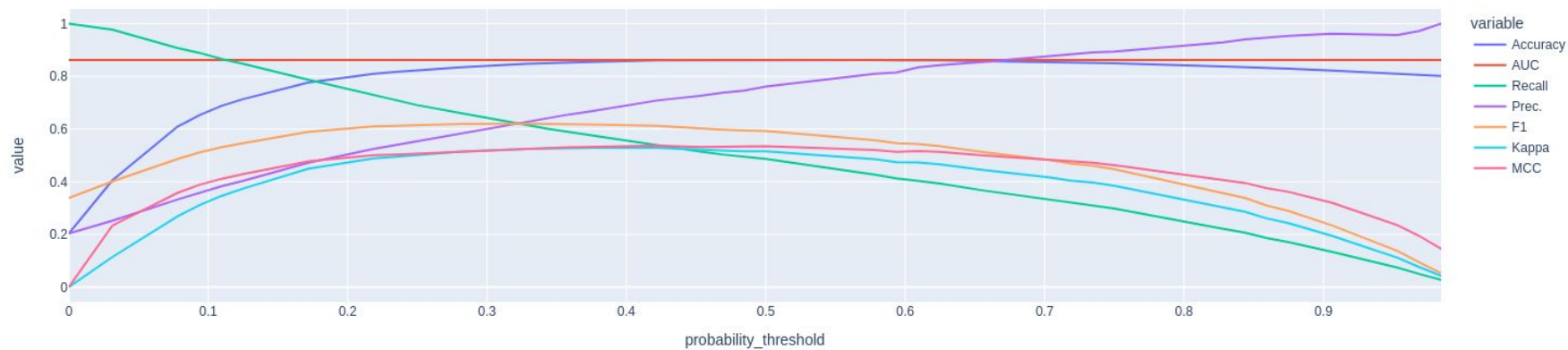
Importancia de características



Optimización



CatBoost Classifier Probability Threshold Optimization (default = 0.5)



Resultados

MCC Inicial → 0.5265

MCC Ajustado → 0.5346

MCC Optimizado → 0.5682

Mejora de 8%

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	threshold
40	CatBoost Classifier	0.8687	0.8787	0.5794	0.721	0.6425	0.5632	0.5682	0.41

MCC Inferencia → 0.5650

Conclusiones y futuras direcciones

- **Datos Demográficos y Financieros Detallados:** Se recomienda la inclusión de información demográfica más detallada sobre los clientes, así como datos financieros adicionales que reflejen su comportamiento profundo con la institución bancaria. Esto podría incluir detalles sobre tendencias de gastos, historial crediticio y productos financieros utilizados.
- **Interacción con Productos:** Para comprender mejor la relación entre los clientes y los productos bancarios, se deberían incorporar datos detallados sobre la interacción del cliente con diversos productos, frecuencia de uso y niveles de satisfacción.
- **Tendencias y Factores Externos:** Explorar cómo las tendencias económicas y otros factores externos podrían influir en el comportamiento de abandono, proporcionando así una visión más completa y contextualizada.