

Tabla de Contenido

1. Antecedentes generales	1
1.1. Sobre el rubro de la inteligencia logística	1
1.2. Sobre la empresa	1
2. Descripción general del proyecto	3
2.1. La problemática y su justificación	3
2.2. Objetivo del trabajo	4
2.2.1. Objetivo general	4
2.2.2. Objetivos específicos	4
2.3. Alcances	5
3. Marco teórico	6
3.1. Fuga de clientes	6
3.2. Modelos de predicción	6
3.2.1. Regresión logística	6
3.2.2. Árbol de decisión	7
3.2.3. XGBoost	7
3.3. Selección de hiperparámetros	7
3.3.1. Grid search	8
3.4. Interpretación de variables	8
3.4.1. Shap analysis	8
3.5. Métricas de desempeño	8
3.5.1. Matriz de confusión	8
3.5.2. Sensibility	9
3.5.3. Specificity	9
3.5.4. Precision	9
3.5.5. Curva ROC	9
3.5.6. AUC (ROC_AUC)	10
4. Metodología	11
4.1. Entendimiento del negocio	11
4.2. Entendimiento de la data	11
4.2.1. Descripción general de la data	11
4.2.2. Variables para modelar	13
4.2.2.1. Variable dependiente	13
4.2.2.2. Variables a nivel cliente y mes	14
4.3. Preparación de la data	17

4.3.1.	Tratamiento inicial de la data	17
4.3.2.	Agrupación y corte de la data	17
4.3.3.	Generación de variables y preparación para modelar	18
4.4.	Modelamiento	18
4.4.1.	XGBoost	19
4.4.2.	Regresión logística	19
4.4.3.	Modelo alternativo: árbol de decisión	19
4.5.	Evaluación	20
4.5.1.	Métricas de desempeño	20
4.5.2.	Interpretabilidad de los modelos	20
5.	Resultados	21
5.1.	Resultado de los modelos	21
5.1.1.	XGBoost	21
5.1.2.	Regresión logística	22
5.2.	Interpretación de variables en modelos	22
5.2.1.	XGBoost	22
5.2.2.	Regresión logística	25
5.3.	Método alternativo: Árbol de decisión simple	25
6.	Discusión	27
7.	Conclusiones y trabajos futuros	29
7.1.	Conclusiones	29
7.2.	Trabajos futuros	30
	Bibliografía	31