

Tabla de Contenido

1. Antecedentes Generales	1
1.1. Características de la Organización	1
1.1.1. Servicios entregados	2
1.1.2. Dimensiones de la Organización	3
1.1.3. Marco Institucional	3
2. Descripción del proyecto y justificación	4
2.1. Área de la organización	4
2.2. Problema u oportunidad	5
3. Objetivos	8
3.1. Objetivo General	8
3.2. Objetivos Específicos	8
4. Marco Conceptual	9
4.1. Ciencia de Datos	9
4.2. Técnicas de Minería de Datos	10
4.2.1. Regresión Logística	10
4.2.2. Árboles de Decisión	10
4.2.3. Random Forest	10
4.2.4. Gradient Boosting Machine	11
4.3. Métricas de Evaluación	11
4.3.1. Matriz de Confusión	11
4.3.2. Precisión	12
4.3.3. Especificidad	12
4.3.4. Sensibilidad (Recall)	12
4.3.5. Exactitud (Accuracy)	12
4.3.6. Curva ROC	12
5. Metodología	13
6. Desarrollo Metodológico	15
6.1. Entendimiento del Negocio	15

6.2.	Entendimiento de la Data	16
6.3.	Preparación de la Data	20
6.3.1.	Incorporación de nuevas variables	22
6.3.2.	Análisis variables seleccionadas	24
6.4.	Modelamiento	35
6.4.1.	Modelos sin datos balanceados	36
6.4.2.	Modelos con datos balanceados	38
6.5.	Evaluación y Comparación de modelos	41
6.6.	Discusión	43
6.6.1.	Análisis monetario	46
6.6.2.	Recomendaciones	46
7.	Conclusiones y Trabajo Futuro	48
7.1.	Trabajo Futuro	49
8.	Bibliografía	51
	Anexo	53