

# TABLA DE CONTENIDO

1.	ANTECEDENTES GENERALES .....	1
1.1.	Antecedentes de la Empresa e Industria .....	1
1.1.1.	Identificación del y descripción del sector industrial .....	1
1.1.2.	Tipo de Organización .....	1
1.1.3.	Servicios y procesos.....	1
1.1.4.	Clientes y usuarios .....	2
1.2.	Descripción del problema y justificación .....	2
1.2.1.	Proceso regular .....	2
1.2.2.	Proceso Demanda anticipada .....	3
1.2.3.	Estimativa valor monetario .....	4
2.	OBJETIVOS .....	6
2.1.	Objetivo general .....	6
2.2.	Objetivos específicos.....	6
3.	ALCANCES .....	7
4.	MARCO CONCEPTUAL .....	8
4.1.	Ciencia de datos.....	8
4.2.	Aprendizaje supervisado .....	8
4.3.	Modelos .....	9
4.3.1.	Árbol de decisión .....	9
4.3.2.	Random Forest.....	9
4.3.3.	Generalized Linear Modeling (GLM) .....	10
4.3.4.	Gradient Boosting Machine (GBM) .....	10
4.3.5.	Redes Neuronales (RN) .....	10

4.4.	Evaluación de los modelos.....	11
4.4.1.	Accuracy .....	11
4.4.2.	Recall.....	11
4.4.3.	Precision.....	11
4.4.4.	F1 Score .....	12
4.4.5.	MCC.....	12
4.4.6.	Curva ROC.....	12
4.5.	Balanceo de datos.....	12
4.5.1.	Sobremuestreo .....	12
4.5.2.	Submuestreo .....	12
5.	METODOLOGIA Y DESARROLLO METODOLOGICO .....	13
5.1.	Entendimiento del negocio .....	13
5.2.	Entendimiento de la data.....	14
5.2.1.	BASE 1 .....	15
5.2.2.	BASE 2 .....	16
5.2.3.	BASE 3.....	16
5.2.4.	Descripción de los datos .....	17
5.3.	Preparación de la data .....	20
5.3.1.	Limpieza .....	20
5.3.2.	Creación de variables.....	21
5.3.3.	Variable respuesta.....	28
5.4.	Modelamiento .....	28
5.4.1.	Variables de los modelos .....	29
5.4.2.	Consideraciones de los datos .....	30

5.4.3.	Particionamiento de los datos .....	31
5.4.4.	Balanceo de datos.....	31
5.4.5.	Modelos .....	31
5.5.	Evaluación y comparación de los modelos .....	34
5.5.1.	Métricas de evaluación.....	34
5.5.2.	Curva ROC y AUC.....	35
5.5.3.	Importancia de las variables.....	36
5.5.4.	Elección del modelo .....	37
6.	CONCLUSIONES .....	38
7.	TRABAJO FUTURO.....	40
8.	BIBLIOGRAFIA .....	41
	ANEXOS.....	43