

# TABLA DE CONTENIDO

I.	INTRODUCCIÓN .....	7
1.1	Empresa.....	7
1.2	<i>Email marketing</i> .....	9
1.3	Justificación del proyecto .....	12
II.	OBJETIVOS .....	16
2.1	Objetivo general .....	16
2.2	Objetivos específicos .....	16
III.	MARCO TEÓRICO .....	17
3.1	Segmentación .....	17
i.	<i>K-medias</i> .....	17
3.2	Definición criterio saturación.....	19
3.2.1	Diferencias en diferencias .....	19
3.2.2	Regresión lineal simple.....	20
3.2.3	<i>Test</i> de hipótesis y p-valor .....	20
3.3	Modelo predictivo.....	21
3.3.1	Selección de variables.....	22
3.3.2	Métrica de ajuste .....	23
3.3.3	Balanceo de datos.....	24
3.3.4	Detección de multicolinealidad .....	24
3.3.5	Regresión logística .....	25
3.3.6	Métrica de desempeño .....	26
IV.	METODOLOGÍA.....	29
4.1	Selección de datos.....	29
4.2	Preprocesamiento .....	29
4.3	Transformación .....	30
4.4	<i>Data mining</i> .....	30
4.4.1	Segmentación de clientes.....	31
4.4.2	Definición de saturación .....	31
4.4.3	Desarrollo de modelo predictivo .....	34
4.4.4	Propuesta de costo de envío .....	35
V.	ALCANCES Y RESULTADOS ESPERADOS .....	36
5.1	Alcances .....	36
5.2	Resultados esperados .....	36
VI.	DESARROLLO METODOLÓGICO.....	37
6.1	Preparación de los datos .....	37
6.1.1	Selección de los datos.....	37
6.1.2	Valores faltantes o erróneos .....	38
6.1.3	<i>Outliers</i> .....	40
6.1.4	Transformación de variables.....	41
6.2	Segmentación de clientes .....	42
6.3	Definición de saturación.....	46
6.3.1	Segmento alto .....	47
6.3.2	Segmento medio y bajo.....	52
6.4	Modelo predictivo de saturación .....	53

6.4.1	Generación de variable dependiente .....	53
6.4.2	Selección de variables con SCAD .....	53
6.4.3	Selección del modelo .....	54
6.4.4	Resolución de potenciales problemas de multicolinealidad .....	55
6.4.5	Desempeño del modelo predictivo .....	57
6.4.6	Calibración del modelo predictivo .....	58
6.5	Propuesta costo de envío .....	59
VII.	CONCLUSIONES .....	61
7.1	Sobre el criterio de saturación .....	61
7.2	Sobre el modelo predictivo .....	62
7.3	Limitaciones y trabajos futuros .....	63
VIII.	GLOSARIO .....	65
IX.	BIBLIOGRAFÍA.....	66
X.	ANEXOS.....	69
10.1	Anexos A: Variables.....	69
10.2	Anexos B: Escenarios sin filtrar transacciones <i>web</i> .....	71
10.3	Anexos C: Selección de variables.....	73
10.4	Anexos D: Selección del modelo .....	79