

Tabla de contenido

1.	Introducción.....	1
1.1	Características de la organización.....	1
1.2	Mercado institucional.....	2
1.3	Desempeño institucional.....	3
2.	Descripción del proyecto y justificación.....	4
2.1	Información del área de la organización.....	4
2.2	Identificación del problema y su relevancia.....	4
2.3	Hipótesis y alternativas de solución para resolver el problema.....	9
3.	Objetivos.....	10
3.1	Objetivo general.....	10
3.2	Objetivos específicos.....	10
4.	Marco conceptual.....	11
4.1	Marco metodológico.....	11
4.1.1	Método RFMC (RFM + C).....	11
4.1.2	Árboles de decisión.....	12
4.1.3	Random forest.....	14
4.1.4	Regresión logística (logit).....	16
4.1.5	Cadenas de markov.....	16
4.1.6	Métricas de desempeño.....	18
4.2	Conceptos propios de la empresa.....	21
4.2.1	Segmentos RFMC.....	21
4.2.2	Macro grupos de taxistas.....	21
5.	Metodología.....	23
5.1	Abstracción del escenario.....	23
5.2	Selección de los datos.....	24
5.3	Limpieza y pre-procesamiento.....	24
5.4	Transformación de los datos.....	24
5.5	Minería de datos.....	24
5.6	Análisis, interpretación y evaluación de resultados.....	25
6.	Alcances.....	26
7.	Resultados esperados.....	26
8.	Desarrollo metodológico.....	27
8.1	Limpieza y pre-procesamiento de los datos.....	27
8.2	Definiciones previas.....	28
8.3	Análisis descriptivo de los datos.....	30
8.3.1	Descripción de transacciones.....	30

8.3.2	Caracterización macro grupos de clientes y variables RFMC.....	33
8.3.3	Caracterización segmentos RFMC	34
8.3.4	Caracterización inactividad.....	37
8.4	Obtención de la muestra	44
8.4.1	Identificación y “adelantamiento” de inactividad	44
8.4.2	Selección de cadenas de actividad	45
8.5	Modelamiento	46
8.5.1	Cadenas de markov.....	46
8.5.1.1	Consideración de últimos movimientos	46
8.5.1.2	Entrenamiento y testeo de modelos	47
8.5.1.3	Resultados.....	48
8.5.2	Árboles de decisión, random forest y logit	50
8.5.2.1	Obtención de variables.....	50
8.5.2.2	Resultados y selección de modelos.....	54
8.5.2.3	Importancia de variables	56
8.5.2.4	Métricas de desempeño	58
8.5.2.5	Análisis de sensibilidad	59
9.	Conclusiones.....	63
10.	Recomendaciones comerciales	65
11.	Trabajo futuro	66
12.	Bibliografía.....	68
13.	Anexos.....	70
13.1	Anexo I: Ecuaciones de Chapman - Kolmogorov.....	70
13.2	Anexo II: Composición de segmentos según puntajes R, F, M, C.....	71
13.3	Anexo III: Análisis continuidad de infrecuencias para los distintos grupos	72
13.4	Anexo IV: Indicadores quincenales de variables R, F, M, C para los distintos grupos	75
13.5	Anexo V: Cantidad y distribución quincenal de taxistas para grupo App y Tarjetas y App.....	81
13.6	Anexo VI: Matrices de correlación de variables para los distintos grupos	83
13.7	Anexo VII: Resultados para definición de parámetro k a utilizar en modelos del grupo Tarjetas.....	84
13.8	Anexo VIII: Resultados para definición de parámetro k a utilizar en modelos del grupo App.....	85
13.9	Anexo IX: Resultados para definición de parámetro k a utilizar en modelos del grupo Tarjetas y App	87
13.10	Anexo X: Análisis de sensibilidad para períodos independientes.....	88